Methods Research Report

Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

Methods Research Report

Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD20850
www.ahrq.gov

Contract No. 290-2007-10055-I

Prepared by:
Tufts Evidence-based Practice Center
Boston, MA

Investigators:
Thomas A. Trikalinos, M.D.
David C. Hoaglin, Ph.D.
Christopher H. Schmid, Ph.D.

AHRQ Publication No. 13-EHC066-EF
March 2013

This report is based on research conducted by the Tufts Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No.290-2007-10055-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report. Authors TAT and CHS are involved in developing open source software for meta-analysis, but this software is available at no charge and the authors receive no financial benefit. All authors have developed multivariate meta-analysis methods.

Suggested citation: Trikalinos TA, Hoaglin DC, Schmid CH. Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes. Methods Research Report. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 13-EHC066-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2013. www.effectivehealthcare.ahrq.gov.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elisabeth U. Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes

Structured Abstract

Background. Many questions in evidence-based medicine involve multiple outcomes. They can be approached with separate, independent meta-analyses, or they can be analyzed jointly, in a single model. We aimed to compare separate (univariate) with joint (multivariate) meta-analysis in real examples and in an illustrative simulation study.

Methods. We screened the whole Cochrane Library of Systematic Reviews (2012, first quarter) to identify sets of univariate meta-analyses of categorical outcomes that can also be analyzed jointly. Eligible were pairs or triplets of meta-analyses comparing the same interventions; having at least seven randomized controlled trials (RCTs) reporting all outcomes; and in which the numbers in the cross-classification of outcomes were exactly recoverable. Examples of outcomes with completely recoverable cross-classification include mutually exclusive outcomes, or sets of outcomes where the one is a subset of the other. We analyzed these data with univariate and multivariate meta-analysis. In an accompanying simulation study, we compared summary estimates and their standard errors with univariate and multivariate meta-analysis.

Results. We identified 45 pairs or triplets of binary meta-analyses corresponding to 1473 RCTs and 258,675 randomized patients. In 38 (of 45) topics the first outcome was a subset of the second outcome; in 5 topics pairs of outcomes were mutually exclusive, and in 2 topics triplets of outcomes had an is-subset-of relationship. The 45 topics pertained to various medical areas (e.g., cardiology, surgery, mental health). Overall, the summary effects for each outcome and the accompanying confidence/credible intervals were very similar with univariate and multivariate meta-analysis (both using the approximate and the discrete likelihood). However, univariate and multivariate approaches yield different confidence/credible intervals for the difference between the summary effects of distinct outcomes (e.g., the difference in the log odds ratio for the first outcome minus the log odds ratio for the second outcome). Depending on the estimated covariance between the compared effects, the multivariate methods can yield tighter or wider confidence intervals than univariate methods. Most likely, systematic review conclusions from the meta-analyses in the empirical sample would remain qualitatively the same with either method of analysis. The simulation analyses were congruent with the aforementioned observations from the empirical analyses.

Conclusions. In the empirical sample and the simulation study, the numerical difference in the summary effects and their confidence intervals between univariate and multivariate meta-analysis was almost always small. In practice, in many (if not most) cases, conclusions based on the main effects of each outcome are likely to remain similar with either method. In principle, multivariate meta-analysis utilizes more information through the correlations; therefore, when possible, it is commendable to use both univariate and multivariate approaches in a sensitivity analysis. Multivariate meta-analysis should be preferred over univariate meta-analysis for estimating differences between outcome-specific summary treatment effects.

Contents

Tables

Figures

Background

The growing number of treatment choices, as well as the rapid escalation in the cost of health care, has spawned a need for scientifically rigorous comparisons of the efficacy and safety of drugs, devices and treatments in clinical practice. To date, most quantitative comparisons carried out by the Evidence Based Practice Centers (EPCs) funded by the Agency for Healthcare Research and Quality (AHRQ) have relied on traditional meta-analysis comparing two treatments with respect to a single outcome. However, many questions involve multiple outcomes. Standard assessments have approached these questions with separate meta-analyses, leading to a plethora of analyses to interpret without any quantitatively rigorous methods for integrating them. As these are multivariate problems, multivariate statistical methods offer a solution.[1-4]

Arguably, joint analysis of all relevant information may be preferable to separate analyses that use only a subset of the available information. For example, imagine that we have *K* studies comparing statins versus no statins for the outcomes of cardiac and noncardiac mortality. The usual approach is to perform two separate meta-analyses: one for cardiac mortality and one for noncardiac mortality. However, both outcomes are evaluated in the same patients (same studies), and are thus stochastically dependent (correlated). Intuitively, knowing something about one outcome tells us something about the other. In the previous example, if too many people die of cardiac causes, fewer people are at risk of dying of noncardiac causes: the proportions dying of the two mutually exclusive causes are negatively correlated. By analyzing the two outcomes jointly we can capitalize on these correlations—an opportunity that is lost with separate, univariate meta-analyses.[2,5]

Statistical methods for simultaneously analyzing multiple outcomes have appeared mainly in the past decade, along with applications to important problems. [2-11] Because they are new, these methods have not migrated into standard clinical research and remain the province of experts. Therefore the empirical evidence base for comparing multivariate meta-analysis of multiple outcomes or multiple follow-ups versus separate univariate meta-analyses is limited. Such empirical data are useful for making informed methodological recommendations regarding the use of multivariate versus univariate meta-analysis in applied systematic review and technology assessment. For example, if the numerical difference in the summary effects and their confidence intervals is always very small, the choice between univariate and multivariate meta-analysis has more academic than practical interest. If nonnegligible differences are common, recommendations can differ.

In this project we perform an empirical evaluation of separate (univariate) versus joint (multivariate) meta-analysis for comparing two treatments with respect to two or more categorical outcomes using real data from the Cochrane Library of Systematic Reviews. We supplement the empirical observations with a simulation study of a large number of scenarios representative of actual analyses.

Project Aims

This project has three aims:

Aim 1: Assemble a database of examples from the Cochrane Library of Systematic Reviews that can be analyzed with multivariate meta-analyses using only information reported in the reviews. (Cochrane systematic reviews perform only univariate meta-analysis.)

Aim 2: Perform an empirical comparison of results with univariate versus multivariate meta-analysis for the database of Aim 1.

Aim 3: Perform illustrative simulation analyses to compare multivariate and separate univariate meta-analyses for the case of two categorical outcomes that have a mutually exclusive relationship, or an is-subset-of relationship.

Aim 1 provides data for the quantitative calculations of Aim 2 and informs the choice of parameters for the simulations of Aim 3.

Methods

Having information on the correlations between the outcomes observed in each study (i.e., within-study correlations) is necessary for the joint meta-analysis of two or more outcomes, unless one is prepared to make assumptions such as that the within and between-study correlations are related.[12]

For some problems, such information is not extractable from data reported in published studies, and one has to obtain individual patient data (for example, see Peter et al.[13]) or to impute the missing information based on prior knowledge, as was done in the example in Berkey et al.[8] (This is true for all joint meta-analyses of continuous outcomes and for many joint meta-analyses between categorical outcomes.) For other problems (all of which pertain to categorical outcomes) the numbers in the cross-classification of outcomes are exactly recoverable. Thus, one can calculate within-study correlations (or equivalent information) from typically reported data. These are cases where:

1. The outcomes are mutually exclusive, that is, a person can experience only one of them. Examples are the pair of outcomes "death from breast cancer" and "death from causes other than breast cancer" and the pair "births by caesarean section" and "spontaneous vaginal births."[5]
2. The people experiencing one outcome are a subset of the people experiencing the other outcome. Such is the pair of outcomes "survival at 6 months" and "survival at 12 months," because those alive at 12 months are a subset of those who were alive at 6 months. Similarly, "withdrawals due to adverse events" are a subset of "withdrawals for any reason."[11]

Table 1 shows an example of mutually exclusive outcomes, and Table 2 an example of outcomes that have an is-subset-of relationship. The examples extend to more than two outcomes in the obvious way.

Table 1. Observed counts for two mutually exclusive outcomes (study k)

Arm	Outcome 1: Breast cancer deaths	Outcome 2: Death from other causes	Alive (remaining)	Total
Comparator (C)	X_{k1}	X_{k2}	X_{k3}	$N_k^{(C)} = X_{k1} + X_{k2} + X_{k3}$
Treatment (T)	Y_{k1}	Y_{k2}	Y_{k3}	$N_k^{(T)} = Y_{k1} + Y_{k2} + Y_{k3}$

Table 2. Observed counts for two outcomes having an is-subset-of relationship (study k)

Arm	Outcome 1: Alive at 12 months	Outcome 2: Alive at 6 months	Total
Comparator (C)	A_{k1}	$A_{k2} (\geq \mathrm{A}_{k1})$	$N_k^{(C)}$
Treatment (T)	B_{k1}	$B_{k2} (\geq B_{k1})$	$N_k^{(T)}$

The outcomes "alive at 12 months" and "alive at 6 months" imply the following three mutually exclusive categories: "alive at 12 months," "dying between 6 and 12 months," and "dying between 0 and 6 months." Denote the counts for the latter three categories in the comparator by F_{k1}, F_{k2}, and F_{k3} and in the treatment by G_{k1}, G_{k2}, and G_{k3} respectively. Then $F_{k1} = A_{k1}$, $F_{k1} + F_{k2} = A_{k2}$, and $F_{k1} + F_{k2} + F_{k3} = N^{(C)}{}_k$, and analogously for the treatment arm.

Aim 1. Formation of the Database

Eligibility Criteria

A single investigator (TT) screened the Cochrane Database of Systematic Reviews, Quarter 1, 2012, to identify pairs or triplets of univariate meta-analyses for which numbers on cross-classification of outcomes are exactly recoverable, and that can be meta-analyzed jointly. Eligible were pairs or triplets of univariate meta-analyses

1. with categorical outcomes;
2. comparing the same interventions;
3. having at least seven randomized controlled trials, RCTs,[a] (or at least half of the RCTs if the number of studies $K>14$) with at least two events and at least 10 patients per arm;
4. where an adequate number of RCTs reported all outcomes of interest.[b] This was defined as at least seven RCTs with two or more events per arm (or at least half of the RCTs if $K>14$); and
5. reporting actual data rather than sensitivity analyses (such as "worst case scenario" analyses).

We excluded Cochrane reviews that have been withdrawn. We did not consider collections of studies that have not been pooled quantitatively by the primary Cochrane reviewers (irrespective of rationale). When the primary Cochrane reviewers performed quantitative syntheses within subgroups of studies but not across these subgroups, we considered each subgroup as a distinct univariate meta-analysis. When the primary reviewers pooled across subgroups, we ignored subgroup classifications and considered the overall synthesis as a single meta-analysis. We excluded meta-analyses of survival outcomes based on approximating a log hazard ratio using the numbers of observed minus expected events.

Some Cochrane reviews could contribute more than one otherwise eligible pair or triplet of outcomes. When applicable, we preferred to form triplets of meta-analyses to selecting one of the three possible pairs in a triplet. To avoid duplication of information, we included only independent pairs or triplets of univariate meta-analyses from a single Cochrane review. These were defined as pairs or triplets pertaining to different intervention comparisons, or pairs or triplets that included nonoverlapping collections of RCTs.

Among nonindependent pairs or triplets of outcomes, we selected the one with the largest number of RCTs reporting all outcomes. We broke any ties by selecting the pair or triplet of meta-analyses with the largest total number of randomized patients, then the largest total number of events, and then randomly.

Screening

We electronically identified all pairs or triplets of meta-analyses with at least seven studies in common. From this set, we manually examined the descriptions of the outcomes and interventions and selected the final candidates based on the full text of the Cochrane review.

[a] Although the vast majority of studies in Cochrane reviews are randomized, a few are not. Nevertheless, we refer to all studies in Cochrane meta-analyses as "RCTs."

[b] Paragraph "On the Number of Parameters to be Estimated" in the Model Fitting subsection below explains the rationale for choosing $K=7$ as the minimum number of studies in a meta-analysis.

Data Extraction

For each eligible pair or triplet of univariate meta-analyses, we recorded the title and identification code of the parent Cochrane review, the compared interventions, outcomes, subgroup definitions (if applicable), the Cochrane Library identification numbers of the included RCTs, and the first author and the year of publication of each RCT. We also recorded the number of events for each outcome and each arm. Finally, we categorized the topics of eligible reviews as pertinent to general medicine, pediatrics, obstetrics and gynecology, nephrology, mental health, cardiology, infectious diseases, and surgery.

Aim 2. Comparison Between Univariate and Multivariate Meta-Analysis

For each eligible pair or triplet of meta-analyses of outcomes we performed univariate (separate) and multivariate (joint) meta-analyses with both fixed and random effects. We used two modeling approaches: the first modeled within-study variability with normal distributions, and the second used discrete distributions (binomial for univariate meta-analysis, multinomial for multivariate). Because most meta-analyses involve clinical and methodological heterogenity, we report results from random effects analyses in the main text of the report. We report results from fixed effects analyses in the Appendix.

Meta-Analysis Models

Assume that we have K studies (indexed by k) reporting M outcomes (indexed by m). The individual outcomes are either mutually exclusive and associated with a single period of follow-up, or have an is-subset-of relationship. Continuing with the example in Table 1, Table 3 shows the true (population) probabilities for the categories defined by outcomes that are mutually exclusive.

Table 3. True probabilities for experiencing two mutually exclusive outcomes (study k)

Arm	Outcome 1: Breast cancer deaths	Outcome 2: Death from other causes	Alive (remaining)
Comparator (C)	$\pi_{k1}^{(C)}$	$\pi_{k2}^{(C)}$	$\pi_{k3}^{(C)} = 1 - \pi_{k1}^{(C)} - \pi_{k2}^{(C)}$
Treatment (T)	$\pi_{k1}^{(T)}$	$\pi_{k2}^{(T)}$	$\pi_{k3}^{(T)} = 1 - \pi_{k1}^{(T)} - \pi_{k2}^{(T)}$

This table corresponds to Table 1 (observed counts).

Sets of outcomes that have an is-subset-of relationship imply a set of mutually exclusive categories (Table 4). The true probabilities for experiencing outcomes that have an is-subset-of relationship are summations over the cells of Table 4, as defined in Table 5. One can extend the notation to more than two outcomes in the obvious way.

Table 4. True probabilities for the mutually exclusive categories implied by two outcomes that have an is-subset-of relationship (study k)

Arm	Alive at 12 months	Dying between 6 and 12 months	Dying between 0 and 6 months
Comparator (C)	$\gamma_{k1}^{(C)}$	$\gamma_{k2}^{(C)}$	$\gamma_{k3}^{(C)} = 1 - \gamma_{k1}^{(C)} - \gamma_{k2}^{(C)}$
Treatment (T)	$\gamma_{k1}^{(T)}$	$\gamma_{k2}^{(T)}$	$\gamma_{k3}^{(T)} = 1 - \gamma_{k1}^{(T)} - \gamma_{k2}^{(T)}$

Table 5. True probabilities for two outcomes that have an is-subset-of relationship (study k) as a function of the mutually exclusive quantities in Table 4

Arm	Outcome 1: Alive at 12 months	Outcome 2: Alive at 6 months
Comparator (C)	$\lambda_{k1}^{(C)} = \gamma_{k1}^{(C)}$	$\lambda_{k2}^{(C)} = \gamma_{k1}^{(C)} + \gamma_{k2}^{(C)}$
Treatment (T)	$\lambda_{k1}^{(T)} = \gamma_{k1}^{(T)}$	$\lambda_{k2}^{(T)} = \gamma_{k1}^{(T)} + \gamma_{k2}^{(T)}$

The γ's are mutually exclusive and exhaustive within each treatment arm (Table 4), but the λ's are not. This table corresponds to Table 2 (observed counts).

Univariate Meta-Analysis

Structural Model

Equal Effects ("Fixed Effects"). For mutually exclusive outcomes, the true effect size for outcome m (here, the log odds ratio, log OR) in study k is

$$\theta_{km} = \text{logit}(\gamma_{km}^{(T)}) - \text{logit}(\gamma_{km}^{(C)}) \quad (1)$$

For outcomes that have an is-subset-of relationship, write

$$\theta_{km} = \text{logit}(\lambda_{km}^{(T)}) - \text{logit}(\lambda_{km}^{(C)}) \quad (2)$$

Under a fixed effects meta-analysis model, the true effect for each outcome is the same in all K studies

$$\theta_{km} = \theta_m . \quad (3)$$

Random-Effects. For random effects we assume that the true effect in each study is normally distributed around a mean θ_m with a between-study variance τ_m^2. Thus (3) is replaced by

$$\theta_{km} \sim \text{N}\ (\theta_m, \tau_m^2) . \quad (4)$$

All other model-defining equations remain the same.

Observational Model

One has two options in modeling within-study variation. The first option is to assume that the sample estimate of the log odds ratio, $\hat{\theta}_{km}$, is normally distributed

$$\hat{\theta}_{km} \sim \text{N}\ (\theta_{km}, \sigma_{km}^2), \quad (5)$$

with the sample conditional variance σ_{km}^2 assumed known and calculated from the data. (This assumption is often made in meta-analysis, but without formal justification.) For mutually exclusive outcomes

$$\hat{\theta}_{km} = \text{logit}\left(\frac{Y_{km}}{N_k^{(T)}}\right) - \text{logit}\left(\frac{X_{km}}{N_k^{(C)}}\right), \text{ and} \quad (6)$$

$$\sigma_{km}^2 = \frac{1}{Y_{km}} + \frac{1}{N_k^{(T)} - Y_{km}} + \frac{1}{X_{km}} + \frac{1}{N_k^{(C)} - X_{km}}. \quad (7)$$

For outcomes with an is-subset-of relationship, substitute A_{km} for X_{km} and B_{km} for Y_{km} in (6) and (7).

Alternatively, one can model the number of events in Table 1 and Table 2 with binomial distributions. For the case of mutually exclusive outcomes

$$X_{km} \sim \mathrm{B}(N_k^{(C)}, \pi_{km}^{(C)}), \text{ and} \quad (8)$$

$$Y_{km} \sim \mathrm{B}(N_k^{(T)}, \pi_{km}^{(T)}). \quad (9)$$

For outcomes with an is-subset-of relationship

$$A_{km} \sim \mathrm{B}(N_k^{(C)}, \lambda_{km}^{(C)}), \text{ and} \quad (10)$$

$$B_{km} \sim \mathrm{B}(N_k^{(T)}, \lambda_{km}^{(T)}). \quad (11)$$

Multivariate Meta-Analysis

We are interested in the same comparisons as in the univariate case. As explained elsewhere, outcomes that are mutually exclusive or have an is-subset-of relationship are correlated and can be analyzed jointly.[5,11] The key is to capitalize on the multinomial structure of the data, which is outlined in Table 3 and in Table 4, respectively.

Structural Model

Equal Effects ("Fixed Effects"). Arrange the M true effects in study k in a column vector $\boldsymbol{\Theta}_k = (\theta_{k1}, \ldots, \theta_{kM})'$, with prime (') denoting transpose and boldface denoting vectors (or matrices). The elements θ_{km} are defined by (1) for mutually exclusive outcomes and (2) for outcomes with an is-subset-of relationship. Under fixed effects, the true effects in all K studies are the same:

$$\boldsymbol{\Theta}_k = \boldsymbol{\Theta}. \quad (12)$$

Random-Effects. In the random effects case, in place of (12) the vector of the true treatment effects $\boldsymbol{\Theta}_k$ is assumed to follow a multivariate normal distribution

$$\boldsymbol{\Theta}_k \sim \mathrm{N}(\boldsymbol{\Theta}, \mathbf{T}), \quad (13)$$

where $\mathbf{T}$ is an $M \times M$ between-study covariance matrix. Trikalinos and Olkin discuss parameterizations for $\mathbf{T}$.[5,11] In this work we use an unstructured $\mathbf{T} = (\tau_{ij})$.

Observational Model

Write $\hat{\boldsymbol{\Theta}}_k = (\hat{\theta}_{k1}, \ldots, \hat{\theta}_{kM})'$ for the estimate of $\boldsymbol{\Theta}_k$. In analogy to the univariate meta-analysis case, one can model within-study variability with a multivariate normal distribution

$$\hat{\boldsymbol{\Theta}}_k \sim \mathrm{N}(\boldsymbol{\Theta}_k, \boldsymbol{\Sigma}_k), \quad (14)$$

where $\boldsymbol{\Sigma}_k$ is an $M \times M$ covariance matrix encoding the correlations between the outcomes within each study. The elements of $\boldsymbol{\Sigma}_k$ are assumed known and are calculated from the data

using the formulas in the Appendix. (As noted in the univariate case, this assumption is not always formally justified.) For details on formulas see the appendices in the papers by Trikalinos and Olkin.[5,11] Note that the formulas for the covariances are different for outcomes that are mutually exclusive compared with outcomes that have an is-subset-of relationship. The analogy between (3) and (12) and between (5) and (14) is obvious.

Alternatively, one can use the multinomial distribution to model observed counts in each arm. For mutually exclusive outcomes, we write using the notation in Table 1 and Table 3:

$$(X_{k1}, \ldots, X_{kM}) \sim \mathrm{M}\left(N_k^{(C)}, (\pi^{(C)}{}_{k1}, \ldots, \pi^{(C)}{}_{kM})\right) \quad (15)$$

for the comparator arm, and

$$(Y_{k1}, \ldots, Y_{kM}) \sim \mathrm{M}\left(N_k^{(T)}, (\pi^{(T)}{}_{k1}, \ldots, \pi^{(T)}{}_{kM})\right) \quad (16)$$

for the treatment arm. For outcomes that have an is-subset-of relationship, we have to model the counts for the mutually exclusive categories implied by the outcomes of interest. As an example, the outcomes "alive at 6 months" and "alive at 12 months" imply the following three mutually exclusive categories: "alive at 12 months," "dying between 6 and 12 months," and "dying between 0 and 6 months." Using the notation in Table 2 and Table 4, we write for the comparator and treatment arms

$$(F_{k1}, F_{k2}, F_{k3}) = (A_{k1}, A_{k2} - A_{k1}, N_k^{(C)} - A_{k2}) \sim \mathrm{M}\left(N_k^{(C)}, (\gamma_{k1}^{(C)}, \gamma_{k2}^{(C)}, \gamma_{k3}^{(C)})\right), \text{ and} \quad (17)$$

$$(G_{k1}, G_{k2}, G_{k3}) = (B_{k1}, B_{k2} - B_{k1}, N_k^{(T)} - B_{k2}) \sim \mathrm{M}\left(N_k^{(T)}, (\gamma_{k1}^{(T)}, \gamma_{k2}^{(T)}, \gamma_{k3}^{(T)})\right), \quad (18)$$

respectively. It is easy to extend (17) and (18) to more than two outcomes with an is-subset-of relationship.

The discrete likelihood methods have several advantages and should be preferred providing software is available to carry them out.[14,15] First, they model the exact probabilistic structure of the data, rather than using large sample approximation as for the normal likelihoods. Second, they require no knowledge of within-study correlation matrices, which may be poorly estimated. Third, they may be estimated without needing to resort to corrections for zero counts where the empirical logits are undefined.

Meta-Analysis Incorporating Missing Outcomes

In many cases, some studies report only a subset of outcomes. Assuming the unreported outcomes are missing at random, the study's contribution to the likelihood can be written in terms of its M_k observed outcomes $\hat{\Theta}_k = (\hat{\theta}_{k1}, \ldots, \hat{\theta}_{kM_k})'$.

Model Fitting

Meta-Analysis Models Using Normal Approximations

Univariate and multivariate meta-analyses using the normal distribution to model within-study variability were fitted with maximum likelihood for fixed effects, and restricted maximum likelihood for random effects. For details see Trikalinos and Olkin.[5,11] For studies with singular covariance matrices (i.e., when the within-study correlation is +1 or -1) we used the ridge regression approach outlined in the appendix of Trikalinos and Olkin.[11] The ridge regression is a form of regularized regression (details in the Appendix). In sensitivity analyses we excluded studies with singular covariance matrices; the results were very similar or identical to those from the ridge regression analyses, and are not shown.

Meta-Analysis Models Using the Discrete Likelihood

Univariate meta-analyses that use the binomial distribution to model counts in each study arm can be fit in the generalized linear mixed models framework using routines readily available in general statistical packages such as *xtmelogit* in Stata or *lmer* in R. However, the multivariate versions of the meta-analysis models specified here cannot be fit in these general routines. The available generalized linear mixed model (GLMM) packages in R, Stata and SAS do not allow the user to specify the random effects distribution in (13), where the random effects pertain to the log ORs. Instead they only allow specification of models in which the random effects are on the logit-transformed probabilities in Table 3 or Table 4.

Optimizing the likelihood for the random effects multinomial model outside a GLMM package is nontrivial, because it involves calculating complicated integrals numerically. Thus we did not develop routines for fitting this model. Instead we fitted the model using Markov Chain Monte Carlo (MCMC) methods in the Bayesian framework. For the Bayesian analyses we used vague (noninformative) prior distributions, as described in the Appendix, as well as prior distributions that were informative on the sign of the between-study correlation coefficients.

On the Number of Parameters to be Estimated

For each univariate meta-analysis the parameters to be estimated are the true mean θ_m for fixed effects, and the true mean and the between-study variance τ_m^2 for random effects. For multivariate meta-analyses, the parameters to be estimated are the M true means (the elements of $\boldsymbol{\Theta}$) for fixed effects, and the $M(M+1)/2$ elements of the unstructured between-study covariance matrix $\mathbf{T}$. To fit the models, we must have more independent data points than parameters. From each study k reporting all outcomes of interest we have $M-1$ data points. So for random effects models, we must have $(M-1)K > M + M(M+1)/2$, or $K > M/(M-1) + M(M+1)/(2(M-1))$. For example, for $M = 2$, $K > 5$. The minimum K is probably not sufficient, and thus we opted to set as an eligibility criterion for this project that meta-analyses should have at least seven studies reporting on all outcomes.

Software

(Restricted) maximum likelihood meta-analyses were performed in Stata using custom code (see the code in Trikalinos and Olkin 2012).[11] Similar analyses can be performed using routines such as *mvmeta* in Stata.[16] Bayesian meta-analyses were performed in JAGS (specifically, using the *rjags* library in R). We used three MCMC chains and a burn-in between 10000 and 50000 iterations. We monitored convergence with the Gelman-Rubin diagnostic for stochastic nodes corresponding to the meta-analysis means and the elements of their between-study covariance matrices. We declared convergence when the 97.5 percentile of the diagnostic was 1.10 or less for all monitored stochastic nodes, provided that on visual inspection the traceplots of the MCMC chains were suggestive of good mixing.

Recorded Results

From each analysis, we recorded the summary effects for each outcome along with the respective variances (and covariances, if applicable), and results on between-study variances (covariances and correlations, as applicable). We also calculated pairwise differences between outcome effects, and the confidence intervals for these differences. These differences are the log

relative odds ratios for observing one outcome versus the other, and are not necessarily helpful or informative for all contexts or topics; however, we calculated them for all topics, because this work makes technical rather than clinical conclusions and observations. We used the *t* distribution with *K-1* degrees of freedom to construct confidence intervals for estimated means and differences of estimated means.

For Bayesian analyses we recorded the median and the 0.025 and 0.975 quantiles of the posterior distribution for each quantity of interest. The latter two can be thought as being the endpoints of a 95% credible interval (95% CrI), the Bayesian analogue of a 95% confidence interval.

Aim 3. Simulation Analysis

We generated illustrative simulation data as described in the next two paragraphs.

Simulation Parameters

Table 6 shows the simulation parameters. These span a representative range of scenarios. Zero between-study standard deviations correspond to fixed-effects realities, and are probably not the norm in real-life applications. However, they are examined for completeness.

Table 6. Simulation parameters

#	Parameter	Values
1	Number of studies per meta-analysis, K_j	10, 20
2	Number of patients per study arm, $N_{j,k}^{(T)} = N_{j,k}^{(C)} = N_j$	50, 100, 500 [Same for all k, and for both study arms]
3	Vector of probabilities for the 3 categorical outcomes in the controls: • For mutually exclusive outcomes: $\boldsymbol{\pi}_{j,k}^{(C)} = \boldsymbol{\pi}_j^{(C)}$, or $(\pi_{j,k1}^{(C)}, \pi_{j,k2}^{(C)}, \pi_{j,k3}^{(C)}) = (\pi_{j,1}^{(C)}, \pi_{j,2}^{(C)}, \pi_{j,3}^{(C)})$ • For outcomes with an is-subset-of relationship: $\boldsymbol{\gamma}_{j,k}^{(C)} = \boldsymbol{\gamma}_j^{(C)}$, or $(\gamma_{j,k1}^{(C)}, \gamma_{j,k2}^{(C)}, \gamma_{j,k3}^{(C)}) = (\gamma_{j,1}^{(C)}, \gamma_{j,2}^{(C)}, \gamma_{j,3}^{(C)})$	(1/6, 1/6, 2/3) [Same for all k]
4	Marginal mean odds ratio for the first outcome, $\theta_{j,1}$	1, 1.5
5	Marginal mean odds ratio for the second outcome, $\theta_{j,2}$	1, 1.5
6	Between-study standard deviation for the log odds ratio of the first outcome, $\tau_{j,1}$	0, 0.1, 0.5
7	Between-study standard deviation for the log odds ratio of the second outcome, $\tau_{j,2}$	0, 0.1, 0.5
8	Between-study correlation between log-odds ratios for the two outcomes, ρ_j	0.25, 0.50, 0.75

j indexes simulation scenarios. Write $\boldsymbol{\theta}_j = (\theta_{j,1}, \theta_{j,2})'$, $\boldsymbol{\tau}_j = (\tau_{j,1}, \tau_{j,2})'$, and $\mathbf{R}_j = \begin{bmatrix} 1 & \rho_j \\ & 1 \end{bmatrix}$.

Generation of Random Data

For each scenario we generated 500 random sets of studies on two correlated binary outcomes. The following pseudo-algorithms in Table 7 (for mutually exclusive outcomes) and Table 8 (for outcomes that have an is-subset-of relationship) explain the generation of simulation data for scenario *j*.

We verified the fidelity of the simulations by comparing the means and covariance matrices of the empirical distributions versus the simulation parameters for the study effects $(\theta_{j,1}, \theta_{j,2})'$ and the proportions of events in the control arms.

Metrics

The main aim of the simulation study is to illustrate comparisons of results from univariate and multivariate meta-analysis. Nevertheless, we calculated the mean squared error (MSE), the bias and the coverage probability for all examined methods (Table 9). We report these metrics for each outcome for analyses using the bivariate random effects method. Additional results are available from the authors.

Table 7. Pseudo-algorithm for generating simulation data for mutually exclusive outcomes

Choose a combination of values for the parameters in Table 6.

Do 500 times (for notational simplicity we drop the index for the simulation draw):

Draw K vectors of counts for the three categories of events in the comparator arms:

$$\mathbf{X}_{j,k} = (X_{j,k1}, X_{j,k2}, X_{j,k3}) \sim \mathrm{M}\left(N_j^{(C)}, (\pi_{j,k1}^{(C)}, \pi_{j,k2}^{(C)}, \pi_{j,k3}^{(C)})\right)$$

Draw *K* vectors of study effects for outcomes 1 and 2

$$\boldsymbol{\theta}_{j,k} = (\theta_{j,k1}, \theta_{j,k2})' \sim \mathrm{N}\left(\boldsymbol{\theta}, \mathbf{T}\right), \text{ where } \mathbf{T} = \mathrm{diag}(\boldsymbol{\tau})\,\mathbf{R}\,\mathrm{diag}(\boldsymbol{\tau})$$

Calculate the probabilities for the three categories of events in the treatment arms:

$$\pi_{j,k1}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k1} + \mathrm{logit}(\pi_{j,k1}^{(C)})\right),$$

$$\pi_{j,k2}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k2} + \mathrm{logit}(\pi_{j,k2}^{(C)})\right), \text{ and}$$

$$\pi_{j,k3}^{(T)} = 1 - \pi_{j,k1}^{(T)} - \pi_{j,k2}^{(T)}.$$

If $\pi_{j,k3}^{(T)} \leq 0$, set

$$\pi_{j,k3}^{(T)} := \frac{10^{-4}}{\pi_{j,k1}^{(T)} + \pi_{j,k2}^{(T)} + 10^{-4}},$$

$$\pi_{j,k2}^{(T)} := \frac{\pi_{j,k2}^{(T)}}{\pi_{j,k1}^{(T)} + \pi_{j,k2}^{(T)} + 10^{-4}}, \text{ and}$$

$$\pi_{j,k1}^{(T)} := \frac{\pi_{j,k1}^{(T)}}{\pi_{j,k1}^{(T)} + \pi_{j,k2}^{(T)} + 10^{-4}}.$$

Draw *K* vectors of counts for the three categories of events in the treatment arms:

$$\mathbf{Y}_{j,k} = (Y_{j,k1}, Y_{j,k2}, Y_{j,k3}) \sim \mathrm{M}\left(N_j^{(T)}, (\pi_{j,k1}^{(T)}, \pi_{j,k2}^{(T)}, \pi_{j,k3}^{(T)})\right)$$

Table 8. Pseudo-algorithm for generating simulation data for outcomes that have an is-subset-of relationship

Choose a combination of values for the parameters in Table 6.

Do 500 times:

Draw *K* vectors of counts for the three mutually exclusive categories implied by the two outcomes (comparator arms):

$$\mathbf{F}_{j,k} = (F_{j,k1}, F_{j,k2}, F_{j,k3}) \sim \mathrm{M}\left(N_j^{(C)}, (\gamma_{j,k1}^{(C)}, \gamma_{j,k2}^{(C)}, \gamma_{j,k3}^{(C)})\right).$$

If those experiencing outcome 1 are a subset of those experiencing outcome 2:

$A_{j,k1} = F_{j,k1}$, and $A_{j,k2} = F_{j,k1} + F_{j,k2}$.

Draw *K* vectors of study effects for outcomes 1 and 2

$\boldsymbol{\theta}_{j,k} = (\theta_{j,k1}, \theta_{j,k2})' \sim \mathrm{N}\left(\boldsymbol{\theta}, \mathbf{T}\right)$, where $\mathbf{T} = \mathrm{diag}(\boldsymbol{\tau})\,\mathbf{R}\,\mathrm{diag}(\boldsymbol{\tau})$

Calculate the probabilities for the three categories of events in the treatment arms:

$$\gamma_{j,k1}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k1} + \mathrm{logit}(\gamma_{j,k1}^{(C)})\right),$$

$$\gamma_{j,k2}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k1} + \mathrm{logit}(\gamma_{j,k1}^{(C)} + \gamma_{j,k2}^{(C)})\right) - \gamma_{j,k1}^{(T)}, \text{ with}$$

$$\gamma_{j,k3}^{(T)} = 1 - \gamma_{j,k1}^{(T)} - \gamma_{j,k2}^{(T)}.$$

If $\gamma_{j,k3}^{(T)} \leq 0$, set

$$\gamma_{j,k3}^{(T)} := \frac{10^{-4}}{\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} + 10^{-4}},$$

$$\gamma_{j,k2}^{(T)} := \frac{\gamma_{j,k2}^{(T)}}{\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} + 10^{-4}}, \text{ and}$$

$$\gamma_{j,k1}^{(T)} := \frac{\gamma_{j,k1}^{(T)}}{\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} + 10^{-4}}.$$

Draw *K* vectors of counts for the three mutually exclusive categories implied by the two outcomes (treatment arms):

$$\mathbf{G}_{j,k} = (G_{j,k1}, G_{j,k2}, G_{j,k3}) \sim \mathrm{M}\left(N_j^{(T)}, (\gamma_{j,k1}^{(T)}, \gamma_{j,k2}^{(T)}, \gamma_{j,k3}^{(T)})\right).$$

Then $B_{j,k1} = G_{j,k1}$, and $B_{j,k2} = G_{j,k1} + G_{j,k2}$.

Table 9. Definition of MSE, bias and coverage probability

Metric	Formula	Description	Comment
Mean squared error (MSE)	$\frac{1}{500}\Sigma_{i=1}^{500}(\hat{\theta}_{j,1i}-\theta_{j,1})^2$ $\frac{1}{500}\Sigma_{i=1}^{500}(\hat{\theta}_{j,2i}-\theta_{j,2})^2$	The average squared difference between the true (simulated) mean and its estimate across the 500 simulation replicates in scenario *j*.	• Desirable to have MSE near zero. • MSE can be high even if bias is 0, because positive and negative deviations of the estimates from the true mean do not cancel out. • MSE is the sum of the variance of an estimate plus the square of its bias.
Bias	$\frac{1}{500}\Sigma_{i=1}^{500}(\hat{\theta}_{j,1i}-\theta_{j,1})$ $\frac{1}{500}\Sigma_{i=1}^{500}(\hat{\theta}_{j,2i}-\theta_{j,2})$	The average difference between the true (simulated) mean and its estimate across the 500 simulation replicates in scenario *j*.	• Desirable to have bias near zero.
Coverage probability	$\frac{1}{500}\Sigma_{i=1}^{500}I(\theta_{j,1}\in[95\%\text{CI of }\hat{\theta}_{j,1i}])$ $\frac{1}{500}\Sigma_{i=1}^{500}I(\theta_{j,2}\in[95\%\text{CI of }\hat{\theta}_{j,2i}])$	The proportion of times the 95% confidence interval of the estimated summary mean contains the true value.	• Desirable to have coverage near 95%. • Coverage higher than 95% indicates an inefficient estimator • Coverage less than 95% indicates an inaccurate estimator

$\hat{\theta}_{j,1i}$ stands for the meta-analysis estimate in draw i, and analogously for the other outcome.

Results

The results are organized by aim.

Description of the Database (Aim 1)

Out of the 4848 reviews in the Cochrane Library, 1919 had at least one collection of studies where the primary Cochrane reviewers performed a meta-analysis of binary data. Ninety-eight reviews identified 1381 candidate pairs of binary meta-analyses with at least seven RCTs reporting on both outcomes and having at least two events and 10 patients per arm. The vast majority were not of the mutually exclusive or is-subset-of type, but rather reported rates of outcomes such as heart attack and stroke without their co-occurrence rate. In such cases, the full combination of outcomes was not exactly recoverable and so did not meet our inclusion criteria. After applying all of the inclusion and exclusion criteria, 29 reviews contributing 45 eligible independent pairs or triplets of binary meta-analyses were eligible: 38 instances pertained to pairs of outcomes where the first outcome was a subset of the second outcome (Table 10, topic numbers 1 through 38); in five instances the pair of outcomes were mutually exclusive (Table 11, topic numbers 39 through 43); and for two topics we identified two triplets of outcomes with an is-subset-of relationship (Table 12, topic numbers 44 and 45). A single Cochrane review on medications for preventing post-operative nausea and vomiting contributed 12 independent pairs of outcomes that have an is-subset-of relationship.

The 45 topics pertain to a variety of clinical questions in cardiology (n=3), general medicine (n=2), pediatrics (n=6), obstetrics and gynecology (n=10), infectious diseases (n=5), nephrology (n=4), mental health (n=2) and surgery (n=13). In total, 1473 RCTs (258,675 randomized patients) were included in the database. The median number of RCTs per topic was 18, ranging from 7 to 203. The median total number of participants per topic was 3733 (range from 304 to 38,923). Twenty one topics had at least 20 RCTs, and 7 topics had at least 50 RCTs. Across the 45 topics, between 62 percent and 100 percent of RCTs reported on all outcomes. In 9 topics all RCTs reported data for all outcomes.

Table 10. Thirty-eight pairs of outcomes having an is-subset-of relationship

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1	O1	O2	O2	O1 & O2	Only O1	Only O2
					ID#*	Description	ID#*	Description	*K* (*N*)	*K* (*N*)	*K* (*N*)
1	Card	CD004587 (1)[17]	Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes	Paclitaxel-eluting stents versus bare metal stents	1, 2	Major adverse cardiac events by 6 months	2, 2	Major adverse cardiac events by 12 months	15 (6325)	6 (1978)	1 (605)
2	Card	CD004587 (1)[17]	Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes	Sirolimus-eluting stents versus bare metal stents	1, 1	Major adverse cardiac events by 6 months	2, 1	Major adverse cardiac events by 12 months	16 (4521)	5 (1141)	4 (1084)
3	Gen	CD002296 (5)[18]	Prevention of NSAID-induced gastroduodenal ulcers	Misoprostol (800 μg/day) versus placebo	8, 3	Dropout due to side events	9, 3	Dropout for any reason	10 (11798)	2 (1095)	1 (277)
4	Ped	CD001145 (6)[19]	Late postnatal corticosteroids for chronic lung disease in preterm infants	Late postnatal corticosteroids versus control	9, 1	Cerebral palsy	11, 1	Death or cerebral palsy	11 (777)	0	0
5	Ped	CD000509 (1)[20]	Inhaled nitric oxide for respiratory failure in preterm infants	Nitric oxide versus control	3, 1	Bronchopulmonary dysplasia among survivors at 36 weeks	4, 1	Death or bronchopulmonary dysplasia at 36 weeks	8 (958)	0	0
6	Ped	CD000551 (3)[21]	Ursodeoxycholic acid for primary biliary cirrhosis	Ursodeoxycholic acid versus nothing	1, 1	Mortality	2, 1	Mortality or liver transplantation	15 (1275)	0	1 (28)
7	Ped	CD001533 (1)[22]	Corticosteroid therapy for nephrotic syndrome in children	Increased versus standard dose of prednisone	4, 3	Number with frequent relapses	2, 3	Number with any relapse	7 (512)	0	0
8	Ped	CD003665 (1)[23]	Vitamin E for prevention of morbidity and mortality in preterm infants	Vitamin E versus placebo/no vitamin E	37, 1	Severe retrolental fibroplasia/retinopathy of prematurity	33, 1	Any retrolental fibroplasia/retinopathy of prematurity	8 (1666)	0	1 (51)
9	Ob Gyn	CD001396 (1)[24]	SSRIs for premenstrual syndrome	SSRIs versus placebo	11, 1	Withdrawal due to adverse events	11, 2	Withdrawal for any reason	32 (3486)	0	5 (247)
10	Ob Gyn	CD001750 (1)[25]	GnRH antagonists for assisted reproductive technology	GnRH antagonist versus long course GnRH agonist	2, 1	Ongoing pregnancy	3, 1	Clinical pregnancy	45 (7209)	4 (638)	17 (2195)
11	Ob Gyn	CD005214 (1)[26]	Depot medroxyprogesterone versus norethisterone oenanthate for long-acting progestogenic contraception.	Depot medroxyprogesterone versus norethisterone oenanthate	1, 1	Discontinuation at 12 months	1, 2	Discontinuation at 24 months	14 (2776)	1 (400)	0
12	Ob Gyn	CD004454 (1)[27]	Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth	Corticosteroids versus placebo/no treatment	6, 1	Neonatal deaths	4, 1	Fetal or neonatal deaths	18 (4140)	5 (513)	0

Table 10. Thirty-eight pairs of outcomes having an is-subset-of relationship (continued)

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1	O1	O2	O2	O1 & O2	Only O1	Only O2
13	Ob Gyn	CD001141 (3)[28]	Support for breastfeeding mothers	All forms of breastfeeding support versus usual care	1, 3	Stopping any breastfeeding by 3 months	1, 5	Stopping any breastfeeding by 6 months	16 (5304)	4 (1247)	2 (664)
14	Ob Gyn	CD001141 (6)[28]	Support for breastfeeding mothers	Professional breastfeeding support versus usual care	1, 3	Stopping any breastfeeding by 3 months	1, 6	Stopping any breastfeeding by 6 months	9 (3539)	1 (507)	1 (849)
15	Ob Gyn	CD000014 (1)[29]	Amnioinfusion for meconium-stained liquor in labour	Amnioinfusion versus no amnioinfusion	6, 1	Caesarean for fetal distress	7, 1	Caesarean overall	11 (3380)	0	3 (615)
16	InfD	CD002848 (2)[30]	Rotavirus vaccine for preventing diarrhoea	Rhesus rotavirus vaccines versus control	2, 1	Severe episodes of rotavirus diarrhoea	1, 1	All episodes of rotavirus diarrhoea	20 (13305)	0	9 (2699)
17	InfD	CD002848 (2)[30]	Rotavirus vaccine for preventing diarrhoea	Bovine rotavirus vaccines versus control	2, 2	Severe episodes of rotavirus diarrhoea	1, 2	All episodes of rotavirus diarrhoea	17 (5283)	0	7 (1640)
18	InfD	CD003774 (1)[31]	Antivirals for preventing CMV disease in solid organ transplant recipients	Antiviral prophylaxis versus placebo/no treatment	1, 2	CMV syndrome	1, 1	All symptomatic CMV disease	19 (1981)	0	8 (411)
19	InfD	CD002898 (1)[32]	Antivirals and other therapies for HSV epithelial keratitis	Acyclovir versus idoxuridine	10, 1	Healing by 7 days	10, 2	Healing by 14 days	9 (476)	0	1 (75)
20	InfD	CD002898 (3)[32]	Antivirals and other therapies for HSV epithelial keratitis	Interferon/nucleoside antiviral versus control	7, 1	Healing by 7 days	7, 2	Healing by 14 days	11 (606)	0	2 (138)
21	Neph	CD003897 (1)[33]	IL2Ra for kidney transplant recipients	IL2Ra versus placebo/no treatment	3, 6	Graft loss censored for death with functioning graft	4, 4	Acute rejection: clinically suspected or biopsy proven	30 (5582)	0	1 (44)
22	Neph	CD003897 (2)[33]	IL2Ra for kidney transplant recipients	IL2Ra versus antithymocyte globulin	3, 3	Graft loss censored for death with functioning graft	2, 3	Graft loss or death with a functioning graft	12 (1394)	0	0
23	Neph	CD003961 (1)[34]	Tacrolimus versus cyclosporin for kidney transplant recipients	Tacrolimus versus cyclosporin	6, 3	Mortality	5, 3	Total graft loss (with death)	14 (2604)	0	0
24	Neph	CD004293 (1)[35]	Immunosuppression for idiopathic membranous nephropathy in adults with nephrotic syndrome	Immunosuppressive therapy versus placebo/no treatment	2, 1	Dialysis or transplantation	3, 1	Dialysis, transplantation or death	10 (620)	0	0
25	MH	CD003197 (3)[36]	Low dosage TCA for depression	Low dosage TCA versus placebo	3, 2	Depression improved by 2 weeks	3, 3	Depression improved by 4 weeks	11 (735)	1 (25)	2 (235)
26	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Ondansetron versus placebo	1, 33	Nausea	3, 28	Nausea or vomiting	132 (16967)	53 (10238)	18 (990)

Table 10. Thirty-eight pairs of outcomes having an is-subset-of relationship (continued)

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1	O1	O2	O2	O1 & O2	Only O1	Only O2
27	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Droperidol versus placebo	1, 17	Nausea	3, 14	Nausea or vomiting	102 (8305)	33 (2199)	22 (1086)
28	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Dexamethasone versus placebo	2, 14	Vomiting	3, 8	Nausea or vomiting	70 (5807)	24 (2276)	4 (173)
29	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Metoclopramide versus placebo	1, 29	Nausea	3, 25	Nausea or vomiting	72 (3685)	18 (978)	14 (684)
30	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Granisetron versus placebo	2, 24	Vomiting	3, 18	Nausea or vomiting	54 (4206)	20 (1554)	2 (93)
31	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Tropisetron versus placebo	4, 39	Rescue antiemetic	3, 40	Nausea or vomiting	28 (2484)	12 (1296)	4 (480)
32	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Dolasetron versus placebo	1, 15	Nausea	3, 12	Nausea or vomiting	14 (2864)	2 (1070)	1 (52)
33	Surg	CD004125 (1)[37]	Prevention of postoperative nausea and vomiting	Hyoscine versus placebo	1, 22	Nausea	3, 20	Nausea or vomiting	15 (1040)	7 (411)	1 (32)
34	Surg	CD004125 (3)[37]	Prevention of postoperative nausea and vomiting	Droperidol versus granisetron	2, 53	Vomiting	3, 42	Nausea or vomiting	24 (1008)	6 (264)	0
35	Surg	CD004125 (3)[37]	Prevention of postoperative nausea and vomiting	Droperidol versus metoclopramide	1, 47	Nausea	3, 46	Nausea or vomiting	32 (964)	9 (209)	5 (238)
36	Surg	CD004125 (3)[37]	Prevention of postoperative nausea and vomiting	Granisetron versus metoclopramide	2, 71	Vomiting	3, 57	Nausea or vomiting	14 (388)	3 (70)	1 (12)
37	Surg	CD004125 (3)[37]	Prevention of postoperative nausea and vomiting	Dexamethasone versus ondansetron	4, 15	Rescue antiemetic	3, 22	Nausea or vomiting	7 (304)	0	0
38	Surg	CD004603 (2)[38]	Perioperative ketamine for acute postoperative pain	Adverse effects	1, 1	Nausea	1, 3	Nausea or vomiting	26 (1283)	0	14 (702)

The number of people experiencing the first outcome is a subset of the number of people experiencing the second outcome.
Card: cardiology; CMV: cytomegalovirus; Gen: general medicine; GnRH: gonadotrophin-releasing hormone; HSV: herpes simplex virus; InfD: infectious disease; MH: mental health; Neph: nephrology; NSAID: nonsteroidal anti-inflammatory drugs; Ob Gyn: obstetrics/gynecology; Ped: pediatrics; SSRI: Selective serotinine reuptake inhibitors; Surg: surgery; TCA: Tricyclic antidepressants. The following are column keys: Comp #: comparison number in the Cochrane review; *K*: number of studies; *N*: number of participants; O1|O2: outcome 1 | 2.
* ID#: The first number is the number of the outcome in the respective Cochrane review; the second number is the number of the subgroup.

Table 11. Five pairs of mutually exclusive outcomes

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1	O1	O2	O2	O1 & O2	Only O1	Only O2
					ID#*	Description	ID#*	Description	*K* (*N*)	*K* (*N*)	*K* (*N*)
39	Gen	CD001431 (1)[39]	Decision aids for people facing health treatment or screening decisions	Decision aids versus usual care	4, 1	Patient controlled decision making	4, 3	Practitioner controlled decision making	11 (1928)	0	1 (171)
40	Ped	CD004000 (1)[40]	Intravenous immunoglobulin for Kawasaki disease in children	Intravenous immunoglobulin versus control	1, 1	Development of chest aortic aneurysms (0 to 30 days)	1, 2	Development of chest aortic aneurysms (31 to 60 days)	10 (1024)	0	0
41	Ob Gyn	CD003766 (3)[41]	Continuous support for women during childbirth	Continuous support versus usual care	3, 1	Spontaneous vaginal birth	4, 1	Caesarean birth	14 (13093)	0	1 (420)
42	Ob Gyn	CD004659 (1)[42]	Antiplatelet agents for primary prevention of preeclampsia and its complications	Antiplatelet versus control	12, 2	Perinatal deaths	12, 1	Stillbirths or miscarriages	31 (24514)	3 (6445)	16 (7964)
43	Ob Gyn	CD004667 (2)[43]	Midwife-led versus other models of care for childbearing women	Midwife-led versus other models of care	9, 2	Spontaneous vaginal birth	7, 2	Caesarean birth	9 (9183)	0	2 (971)

Outcomes 1 and 2 are mutually exclusive. Gen: general medicine; Ob Gyn: obstetrics/gynecology; Ped: pediatrics.
The following are column keys: Comp #: comparison number in the Cochrane review; *K*: number of studies; *N*: number of participants; O1|O2: outcome 1 | 2.
*ID#: The first number is the number of the outcome in the respective Cochrane review; the second number is the number of the subgroup.

Table 12. Two triplets of outcomes having an is-subset-of relationship

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1	O1	O2	O2	O3	O3
					ID#*	Description	ID#*	Description	ID#*	Description
44	MH	CD001026 (1)[44]	Antidepressants plus benzodiazepines for major depression	Antidepressants plus benzodiazepines versus antidepressants	7, 1	Depression improved by 1 week	7, 2	Depression improved by 2 weeks	7, 3	Depression improved by 4 weeks
45	Card	CD002230 (5)[45]	Phosphodiesterase III inhibitors for heart failure	Phosphodiesterase III inhibitors versus control	3, 1	Sudden death	2, 1	Cardiac death	1, 1	Total mortality

The number of people experiencing the first outcome is a subset of the number of people experiencing the second outcome, which in turn is a subset of those experiencing the third outcome. MH: mental health; Card: cardiology.
The following are column keys: Comp #: comparison number in the Cochrane review; O1|O2|O3: outcome 1 | 2 | 3.
*ID#: The first number is the number of the outcome in the respective Cochrane review; the second number is the number of the subgroup.
For the first topic (number 44) a total of 10 RCTs with a total of 599 people provided data on all three outcomes. For the second topic (number 45), 13 RCTs (7337 people) provided data for all three outcomes, two RCTs (469 people) provided data only for the first and the third outcome, one RCT (230 people) provided data only for the second and the third outcome, and three RCTs (262 people) provided data only for the third outcome.

Empirical Comparisons (Aim 2)

We first describe results with meta-analysis models that use the normal approximation. Subsequently, we present the corresponding analyses with models that use the binomial or multinomial distribution to model the underlying data.

Normal Approximation Methods

Corrections for Singularity in Covariance Matrices

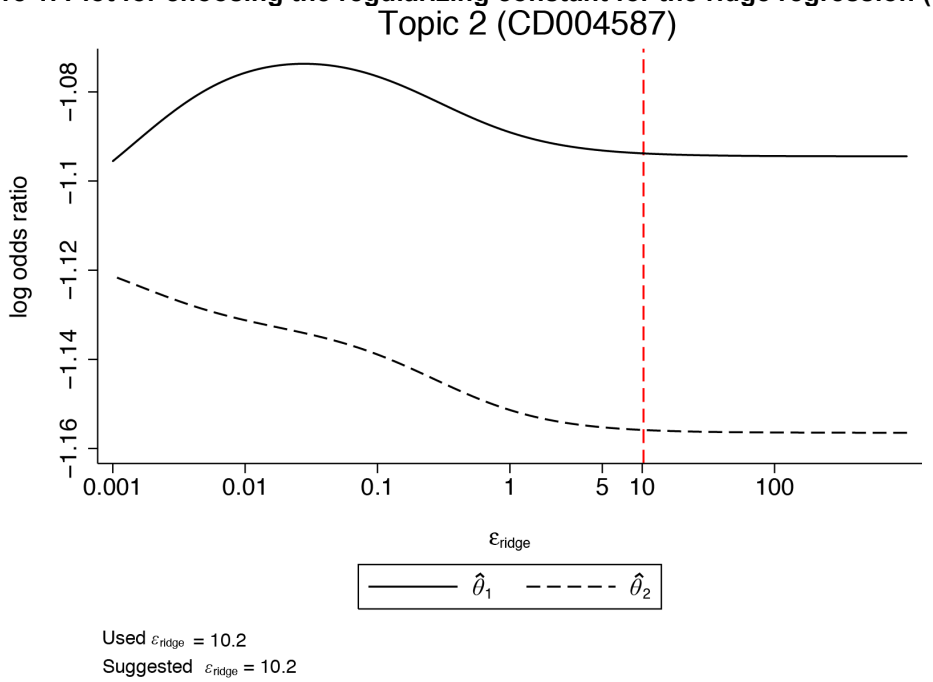
Eighty studies in 21 topics, all in outcomes having an is-subset-of-relationship, had singular covariance matrices. Based on the formulas in the Appendix, it is obvious that singular covariance matrices can arise when the number of people experiencing the first outcome is exactly the same as the number experiencing the second outcome. For example, in topic 2, comparison between sirolimus-eluting and bare metal stents for acute cardiac disease, in the STRATEGY trial, the number of major cardiac events was the same at 6 months and 12 months, that is, no additional events were observed between 6 and 12 months.[46,47] The multivariate analyses used a ridge regression approach when at least one study had a singular covariance matrix. Table 13 shows the ridge regression factors (regularizing constants or regularizers) used. Using larger values or excluding the studies with singular covariance matrices from the analysis yields very similar results to those presented here (not shown). Figure 1 shows a plot of the meta-analysis means versus a range of values for the regularizing constant $\varepsilon_{\text{ridge}}$ for topic 2.

Table 13. Topics with RCTs with singular covariance matrices

Topic	Number of RCTs with singular matrices	Number of RCTs reporting two or more outcomes	Total number of RCTs	$\varepsilon_{\text{ridge}}$
2	1	16	25	10.2
3	1	10	13	15.8
4	1	11	11	2.5
6	5	15	16	4.8
9	2	32	37	2.5
10	1	45	66	23.4
12	1	18	23	6.2
22	2	12	12	4.3
23	1	14	14	2.7
24	4	10	10	4.4
26	18	132	203	70.8
27	14	102	157	77.6
28	1	70	98	7.6
29	10	72	104	32.4
31	1	28	44	4.1
32	3	14	17	10.0
33	1	15	23	3.2
34	1	24	30	3.4
35	4	32	46	6.5
38	7	26	40	14.8
45	1	16	17	575.4

The correction factors used in the ridge regression approach were selected using an objective heuristic (described in the Appendix). The remaining topics did not have any RCTs with singular covariance matrices.

Figure 1. Plot for choosing the regularizing constant for the ridge regression (example, topic 2)

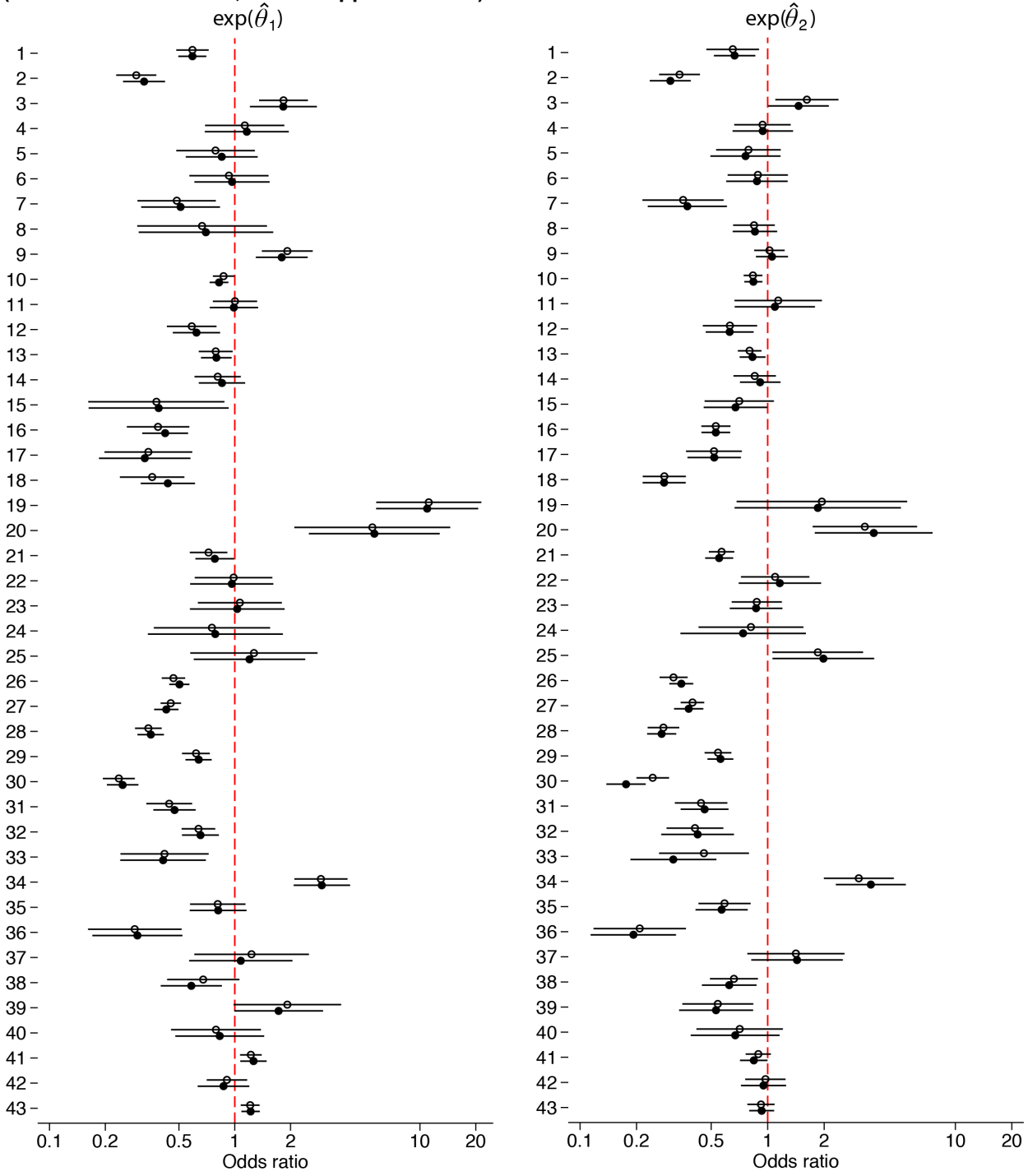


The suggested value of the regularizing constant uses a heuristic that examines the smoothed numerical derivatives with respect to $\varepsilon_{\text{ridge}}$ of the meta-analysis means and of the trace and determinant of the inverse covariance matrix for the means. An investigator examined such plots for all topics listed in Table 13; in all cases the $\varepsilon_{\text{ridge}}$ value suggested by the heuristic was deemed acceptable. An infinitely large value for $\varepsilon_{\text{ridge}}$ is equivalent to excluding studies with singular covariance matrices from the multivariate meta-analysis.

Figure 2 shows summary results with univariate and multivariate (bivariate) meta-analyses using the normal likelihood (in Appendix Table 1 we give the same results in tabular form). Figure 3 shows the respective results for topics 44 and 45, which have three outcomes with an is-subset-of relationship (the data in the figure are reported in tabular form in Appendix Table 2). Generally, the point estimates and the lengths of the confidence intervals are comparable using the two methodologies.

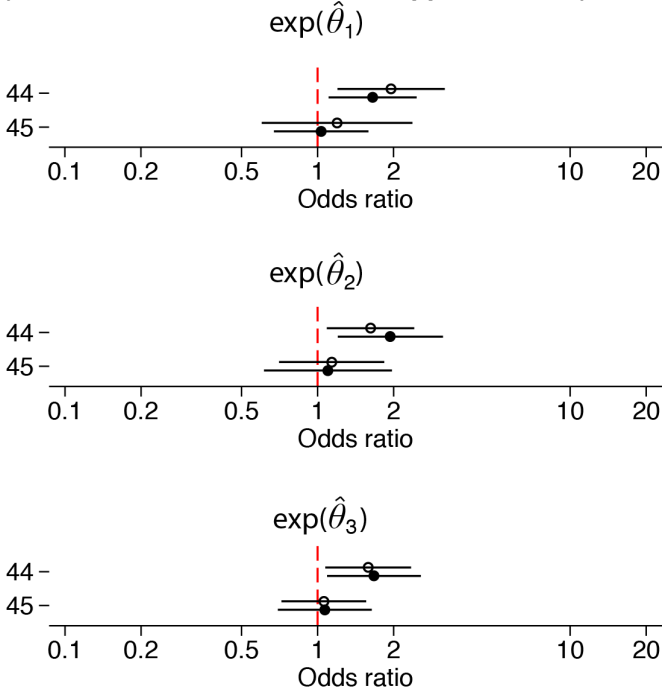
Some exceptions are notable. The most prominent is in topic 30, comparison of granisetron versus placebo for postoperative nausea and vomiting, where the summary odds ratio for "nausea or vomiting" (the second outcome) is more extreme in bivariate meta-analysis (0.18 95% CI 0.14 to 0.22) than in univariate meta-analysis (0.24, 95% CI 0.20 to 0.30). The summary odds ratios for the first outcome ("vomiting") are 0.25 (0.20 to 0.30) and 0.24 (0.19 to 0.30), respectively. Nevertheless, even this numerical difference would not change conclusions on the effectiveness of granisetron. Topic 10 differs qualitatively for outcome 1, in that one of the confidence intervals includes the null value of 1 whereas the other does not. A similar comment applies to topics 15 and 41 for outcome 2. However, the actual numerical differences are very small. In all likelihood, systematic review conclusions that are based on the meta-analyses examined here would be extremely similar with either method of analysis.

Figure 2. Summary estimates from univariate and bivariate meta-analysis in topics 1 through 43 (REML random effects, normal approximation)



Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$: meta-analysis mean for the first outcome; $\hat{\theta}_2$: meta-analysis mean for the second outcome. For topics 1 through 38 those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive.

Figure 3. Summary estimates from univariate and trivariate meta-analysis in topics 44 and 45 (REML random effects, normal approximation)



Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$ through $\hat{\theta}_3$: meta-analysis means for the first through third outcome.

As evident from Figure 2, Figure 3 and the Appendix tables, multivariate and univariate analyses yielded confidence intervals of different length. Across all 45 topics, the ratio of the standard errors with multivariate versus univariate analyses ranged between 0.6 and 1.4 for the first outcome (median ratio 0.99) and between 0.8 and 1.3 for the second outcome (median ratio 1.00). The ratio was 1.1 for the third outcome in topics 44 and 45. Compared with univariate analyses, multivariate analyses yielded smaller standard errors (ratio of standard errors less than 1) in 25 and 22 topics for the first and second outcome, respectively.

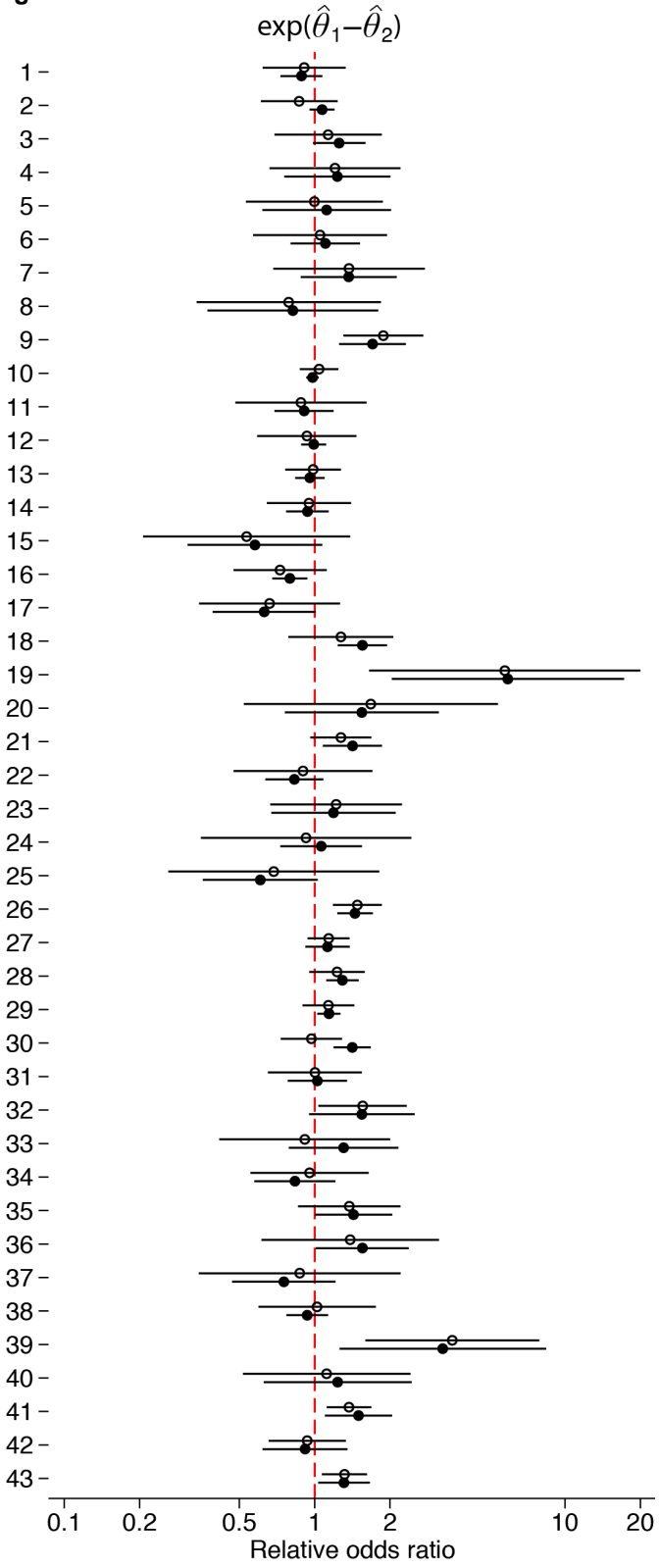
Between-Outcome Relative Odds Ratios (Differences in Treatment Effects Between Outcomes)

It is instructive to calculate differences in the meta-analysis means. As discussed in the Methods section, such differences are log relative odds ratios for experiencing the first versus the second outcome in the treatment versus the comparator group. Depending on the context of the topic and the definitions of the outcomes, these differences may or may not be helpful. For example, if the first outcome is "alive at 12 months" and the second outcome is "alive at 6 months," the difference in the outcomes can inform on whether the treatment log odds ratio differs between earlier and later time points. However, it is unclear whether the relative odds ratio is helpful or informative for pairs of outcomes such as "spontaneous vaginal birth" and "caesarean birth." Regardless, forming differences between the outcomes presents an opportunity for technical observations and comments.

Figure 4 shows the relative treatment effects for univariate and bivariate meta-analyses in topics 1 through 43. Figure 5 shows the corresponding results for topics 44 and 45. The tabular representations are in Appendix Tables 1 and 2. Univariate and multivariate meta-analyses yield confidence intervals of different lengths. The reason is that univariate meta-analysis ignores the

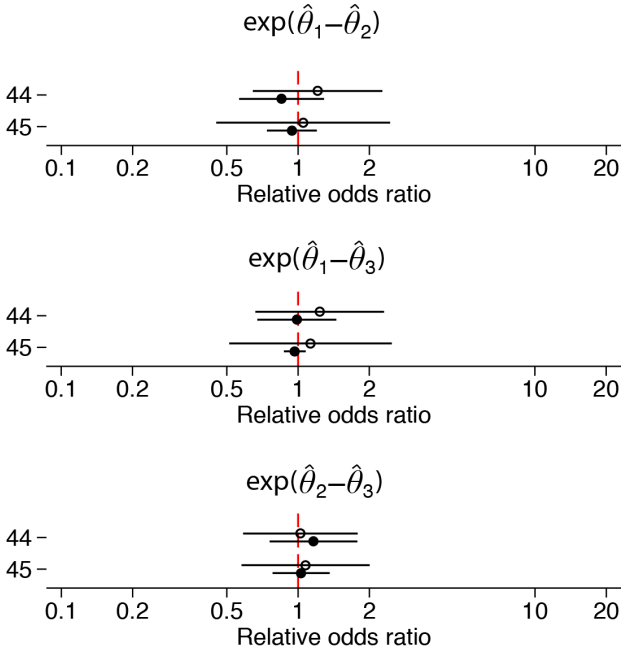
correlation between the meta-analysis effects, whereas multivariate meta-analysis does not. When this correlation is positive, the standard error of the relative effect tends to be smaller in multivariate than in univariate meta-analysis; when the correlation is negative (as can happen for mutually exclusive outcomes), standard errors derived from multivariate analyses tend to be larger than those derived from univariate analyses (Figure 6). These differences can lead to some quantitatively different conclusions as with topic 18, in which the much smaller bivariate model standard error leads to a significant difference in the relative odds ratio not seen in the univariate analysis, or topic 2, in which the odds ratio is larger for the second outcome in the univariate analysis, but larger for the first outcome in the multivariate one.

Figure 4. Relative summary odds ratios from univariate and bivariate meta-analyses in topics 1 through 43



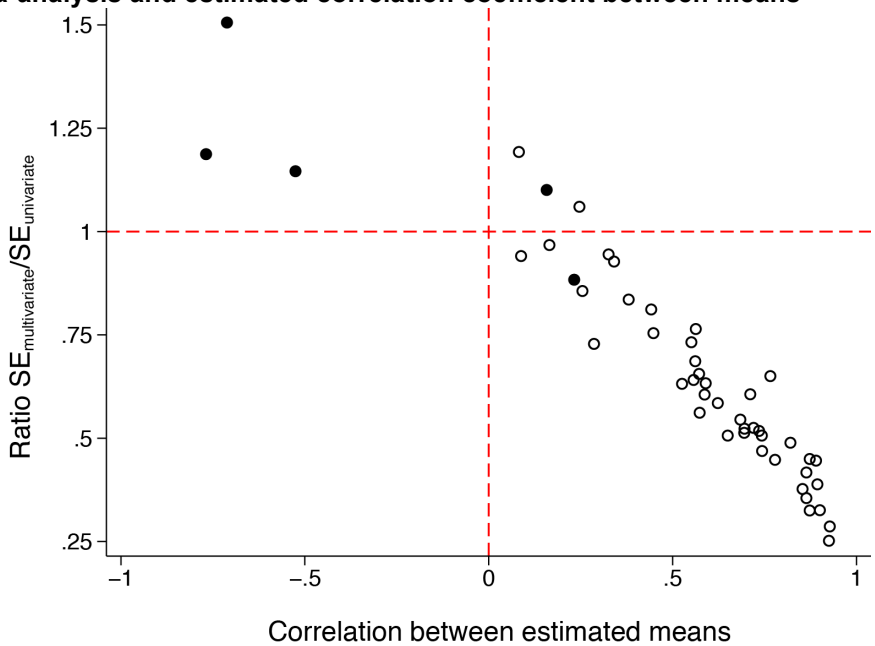
Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$ and $\hat{\theta}_2$: meta-analysis means for the first and second outcome, respectively.

Figure 5. Treatment effects with univariate and trivariate meta-analysis in topics 44 and 45.



Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$ through $\hat{\theta}_3$: meta-analysis means for the first through third outcome.

Figure 6. Ratio of standard errors for log relative odds ratios from multivariate and univariate meta-analysis and estimated correlation coefficient between means

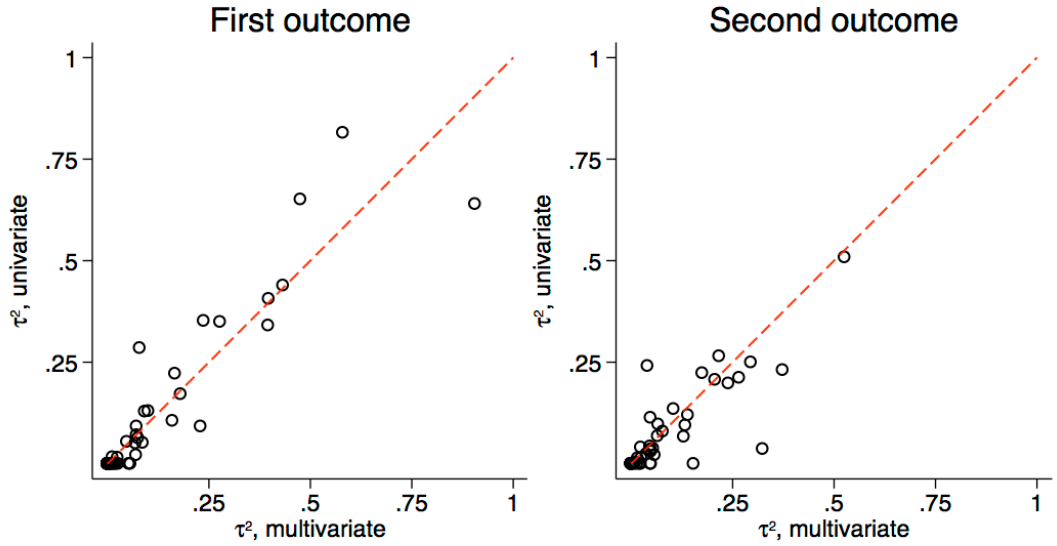


Filled circles denote topics where the jointly meta-analyzed outcomes are mutually exclusive. Empty circles denote topics where outcomes have an is-subset-of relationship.

Comparison of Between-Study Variances With Univariate Versus Multivariate Analyses

Figure 7 shows estimates of between-study variances with univariate versus multivariate analyses. Between-study variance estimates were larger with multivariate analyses compared with univariate in 32 out of 45 examples for the first outcome, 36 out of 45 for the second outcome and for both examples for the third outcome (sign test $p<0.10^{-6}$ that variances are larger with multivariate analyses across all three outcomes). The above is consistent with the observations of Riley et al. who conjecture that the difficulty in estimating the between-study correlation may result in slightly inflated estimates for between-study variance with multivariate versus univariate meta-analysis.[12,48]

Figure 7. Between-study variance estimates in univariate versus multivariate analyses (frequentist analyses)



Information Flow Across Outcomes in Multivariate Meta-Analysis

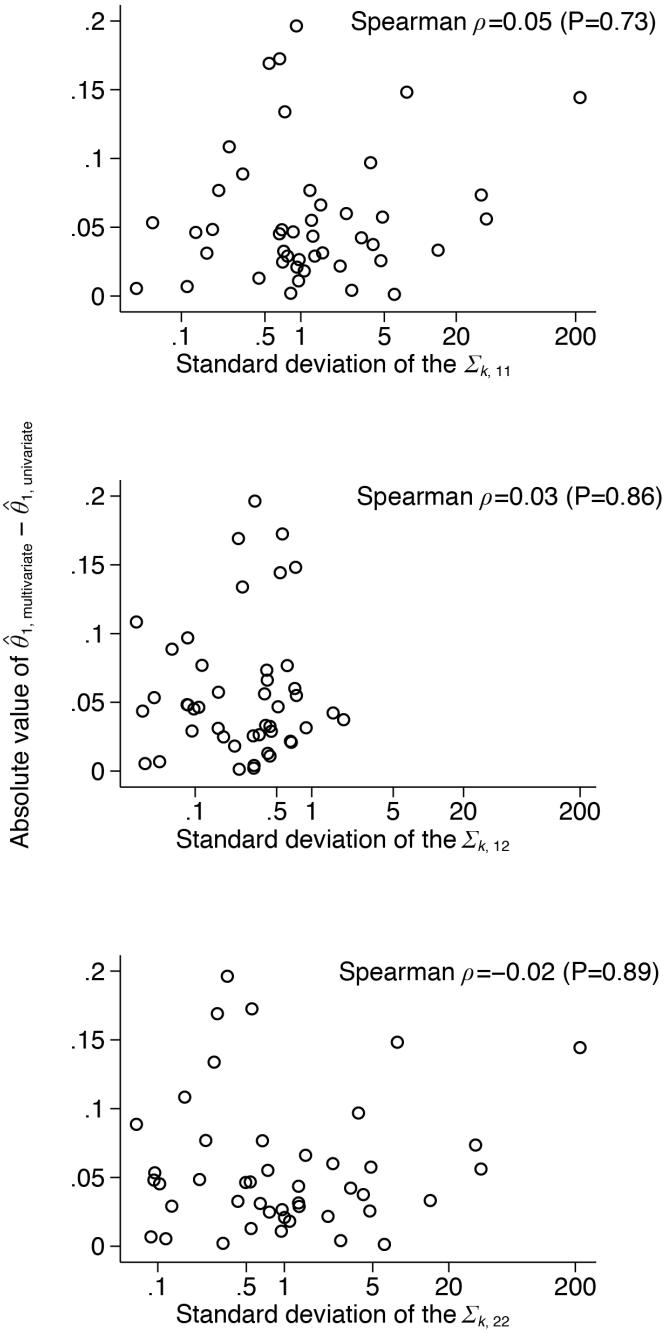
Riley et al. provide formulas for the point estimates of the means and the standard errors for random effects bivariate meta-analysis in a restricted iterative generalized least squares (RIGLS) framework.[48] Upon inspection of the formulas it is evident that when the two outcomes are analyzed jointly, the point estimate of the first outcome depends on the study results for the second outcome if the following three conditions hold: (a) the within-study covariances $\Sigma_{k,12}$ are nonzero; (b) the within-study variances $\Sigma_{k,11}$ are not all equal ($\Sigma_{k,11} \neq \Sigma_{11}$); and (c) the within-study covariances $\Sigma_{k,12}$ are not all equal ($\Sigma_{k,12} \neq \Sigma_{12}$). Similarly, the point estimate of the second outcome depends on the study results for the first outcome if a set of analogous conditions holds. (The only change is in condition (b), which becomes $\Sigma_{k,22} \neq \Sigma_{22}$). The covariance matrices for the summary effects always depend on both outcomes. Riley et al. suggest that information flow will increase as the elements of the within-study covariance matrices become more dissimilar.

If information flows from the second outcome to the first outcome, the point estimates of the first outcome may be different under the univariate and multivariate meta-analysis. Reciprocally, if information flows from the first outcome to the second outcome, the point estimates of the

second outcome may be different with univariate and multivariate meta-analysis. Figure 8 and Figure 9 examine the relationship between the absolute difference in the meta-analysis estimates with univariate and multivariate methods, and the standard deviation of the within-study variances and covariances across the 45 topics. Under the Riley hypothesis, one might expect that the absolute differences would increase with these standard deviations. However, all the correlations are very small so in this sample of 45 topics, we see no strong evidence of information flow.

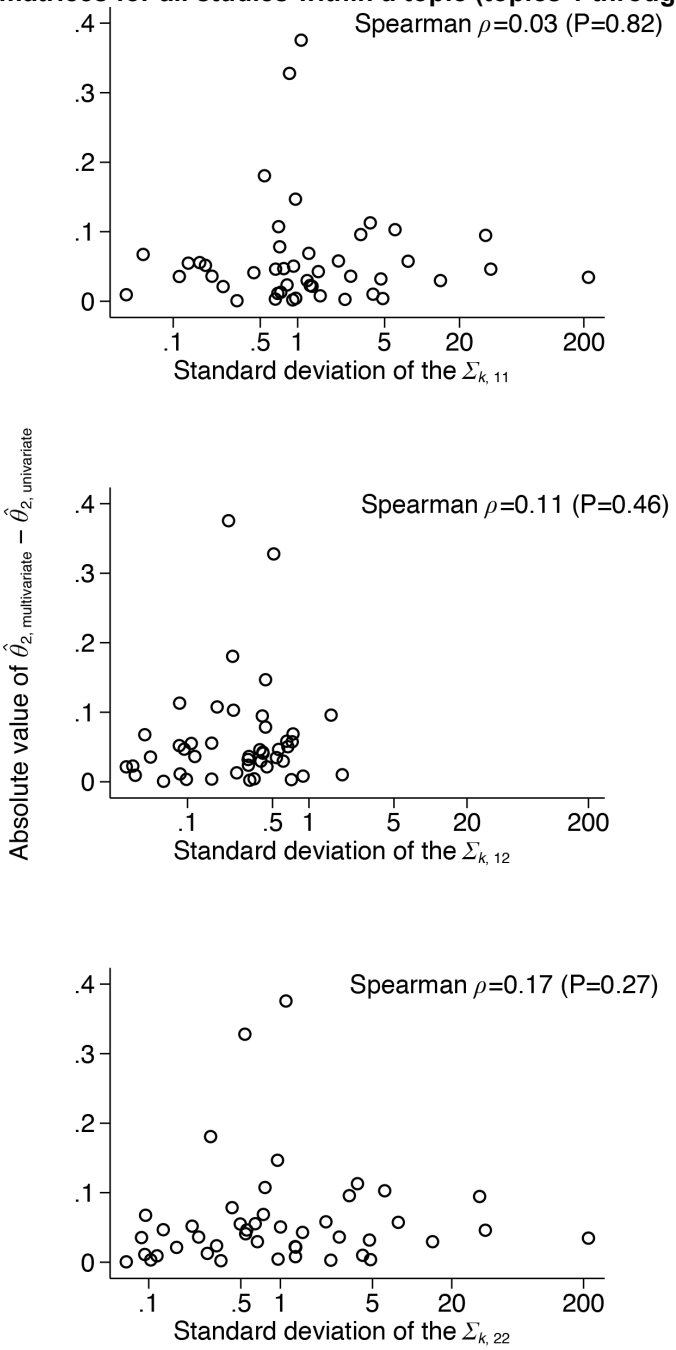
Finally, information can flow towards the first meta-analysis mean, when additional studies exist for the second outcome, and reciprocally. Figure 10 provides scatter plots of the absolute difference in the meta-analysis estimates with multivariate and univariate models and the number of studies reporting only the other outcome. Overall, the observed association is not beyond what is expected by chance.

Figure 8. Difference in the point estimate for the first outcome with univariate and multivariate meta-analysis versus the standard deviation of the elements of the within-study covariance matrices for all studies within a topic (topics 1 through 45)



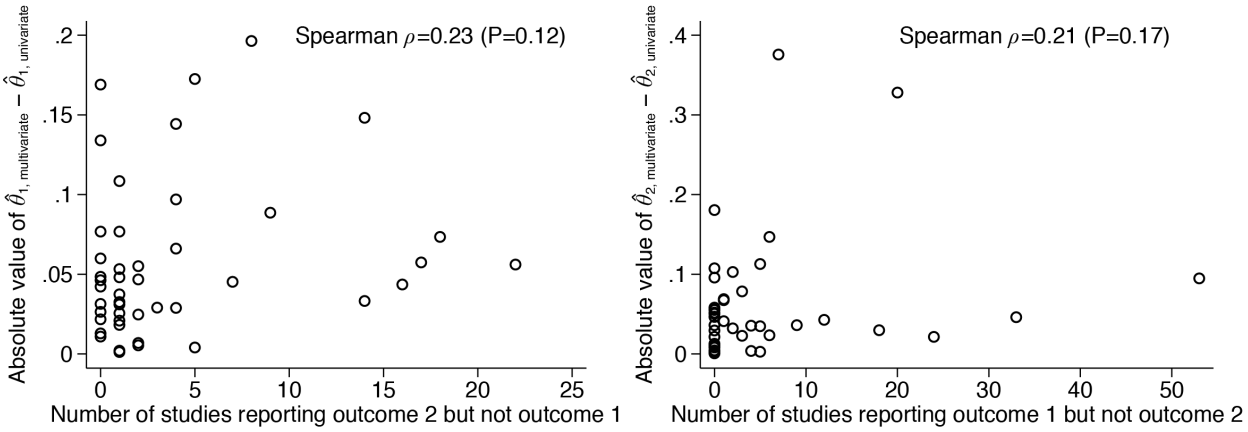
Differences between univariate and multivariate point estimates for the first outcome suggest flow of information from the second outcome to the first outcome.

Figure 9. Difference in the point estimate for the second outcome with univariate and multivariate meta-analysis versus the standard deviation of the elements of the within-study covariance matrices for all studies within a topic (topics 1 through 45)



Differences between univariate and multivariate point estimates for the second outcome suggest flow of information from the first outcome to the second outcome.

Figure 10. Difference in the point estimates with univariate and multivariate meta-analysis versus the number of studies reporting the other outcome (topics 1 through 45)



Estimation of the Between-Study Correlation Parameters

Between-study correlations were poorly estimated in almost all topics (Table 14). The point estimates are approximately +1 or –1; the respective variances (obtained by applying the delta method to the fitted multivariate meta-analysis model results) were very large, indicating a relatively flat likelihood with respect to the correlation parameter. In some sense, the fitted models provide very little information about the correlation. This observation is concordant with published analyses of single examples.[11]

Consequently, it is not surprising that the estimated means from the multivariate models differ little from their univariate counterparts. In general, one should not expect much information flow in three situations: 1) correlations are not well-estimated; 2) correlation is zero (because then the outcomes are independent); and 3) correlations are +1 or –1 (because then the outcomes are redundant).

Table 14. Estimated between-study correlation between the first and second outcome

Topic	$\hat{\rho}_{12}$ var($\hat{\rho}_{12}$)	Topic	$\hat{\rho}_{12}$ var($\hat{\rho}_{12}$)	Topic	$\hat{\rho}_{12}$ var($\hat{\rho}_{12}$)
1	1.00 (-)	16	1.00 (-)	31	1.00 (-)
2	1.00 (-)	17	0.63 (0.10)	32	-1.00 (-)
3	1.00 (-)	18	-0.78 (-)	33	1.00 (-)
4	-1.00 (-)	19	1.00 (-)	34	1.00 (-)
5	-1.00 (-)	20	1.00 (-)	35	0.08 (-)
6	1.00 (-)	21	-1.00 (-)	36	0.05 (-)
7	1.00 (-)	22	1.00 (-)	37	0.99 (-)
8	1.00 (-)	23	-0.43 (2.59)	38	-1.00 (-)
9	-1.00 (-)	24	1.00 (-)	39	-1.00 (-)
10	-1.00 (0.04)	25	0.91 (0.13)	40	1.00 (-)
11	1.00 (-)	26	0.13 (0.95)	41	-1.00 (-)
12	1.00 (-)	27	1.00 (-)	42	1.00 (-)
13	1.00 (-)	28	1.00 (-)	43	-0.99 (-)

Table 14. Estimated between-study correlation between the first and second outcome (continued)

Topic	$\hat{\rho}_{12}$ var($\hat{\rho}_{12}$)	Topic	$\hat{\rho}_{12}$ var($\hat{\rho}_{12}$)	Topic	$\hat{\rho}_{12}$ var($\hat{\rho}_{12}$)
14	1.00 (-)	29	0.21 (-)	44	-1.00 (-)
15	1.00 (-)	30	1.00 (-)	45	1.00 (-)

The dash ("-") denotes an estimated variance larger than 1000. In the trivariate meta-analyses in topics 44 and 45, the estimated correlation coefficients (and variances) between the first and third outcome were -1 (42.2) and 0.98 (0.02), respectively; and between the second and third outcome they were 1 (>1000) and 0.97 (>1000).

Meta-Analyses Using the Binomial or Multinomial Likelihood

Analyses with these models were performed in the Bayesian framework (see "Meta-Analysis Models Using the Discrete Likelihood," above). Modeling within-study variation based on the binomial (for univariate meta-analyses) or the multinomial (for multivariate meta-analyses) distribution does not require continuity corrections or the use of a regularizer (a ridge regression correction). Overall, results were very similar to those obtained from models using the normal approximation. For example, Figure 11 and Figure 12 are the counterparts of Figure 2 and Figure 3, respectively. Similarly, Figure 13 and Figure 14 are the counterparts of Figure 4 and Figure 5, and Figure 15 is the counterpart of Figure 7. The numerical results are provided in the Appendix Tables 5 and 6.

The posterior distribution of the between-study correlation parameter(s) reflected the prior chosen. With the noninformative prior, the posterior medians were closer to zero with extremely wide credible intervals that almost covered the entire potential range from -1 to 1. The prior informative on the sign moved the posterior median toward that boundary and slightly tightened the credible interval, which still remained wide, however (Table 15). This is congruent with the findings in Table 14 for analyses using the normal approximation. The signs of the correlation coefficients from analyses using the normal approximation (frequentist analyses) and analyses using the discrete likelihood (Bayesian analyses, noninformative priors) generally agreed (κ = 0.42, P=0.001).

Table 15. Posterior median and 95% credible interval for between-study correlations with uninformative priors and priors informative on the correlation sign

Quantity or topic	Noninformative prior	Prior informative on the sign of the correlation
ρ_{12}		
1	0.42 (-0.82, 0.96)	0.69 (-0.66, 0.98)
2	0.45 (-0.88, 0.98)	0.82 (-0.58, 0.99)
3	0.29 (-0.91, 0.97)	0.68 (-0.73, 0.98)
4	0.28 (-0.86, 0.96)	0.64 (-0.75, 0.98)
5	0.08 (-0.93, 0.94)	0.48 (-0.84, 0.98)
6	0.76 (-0.68, 0.99)	0.90 (-0.26, 0.99)
7	0.19 (-0.92, 0.96)	0.66 (-0.78, 0.99)
8	0.04 (-0.94, 0.94)	0.57 (-0.80, 0.99)
9	-0.03 (-0.95, 0.92)	0.32 (-0.86, 0.96)

Table 15. Posterior median and 95% credible interval for between-study correlations with uninformative priors and priors informative on the correlation sign (continued)

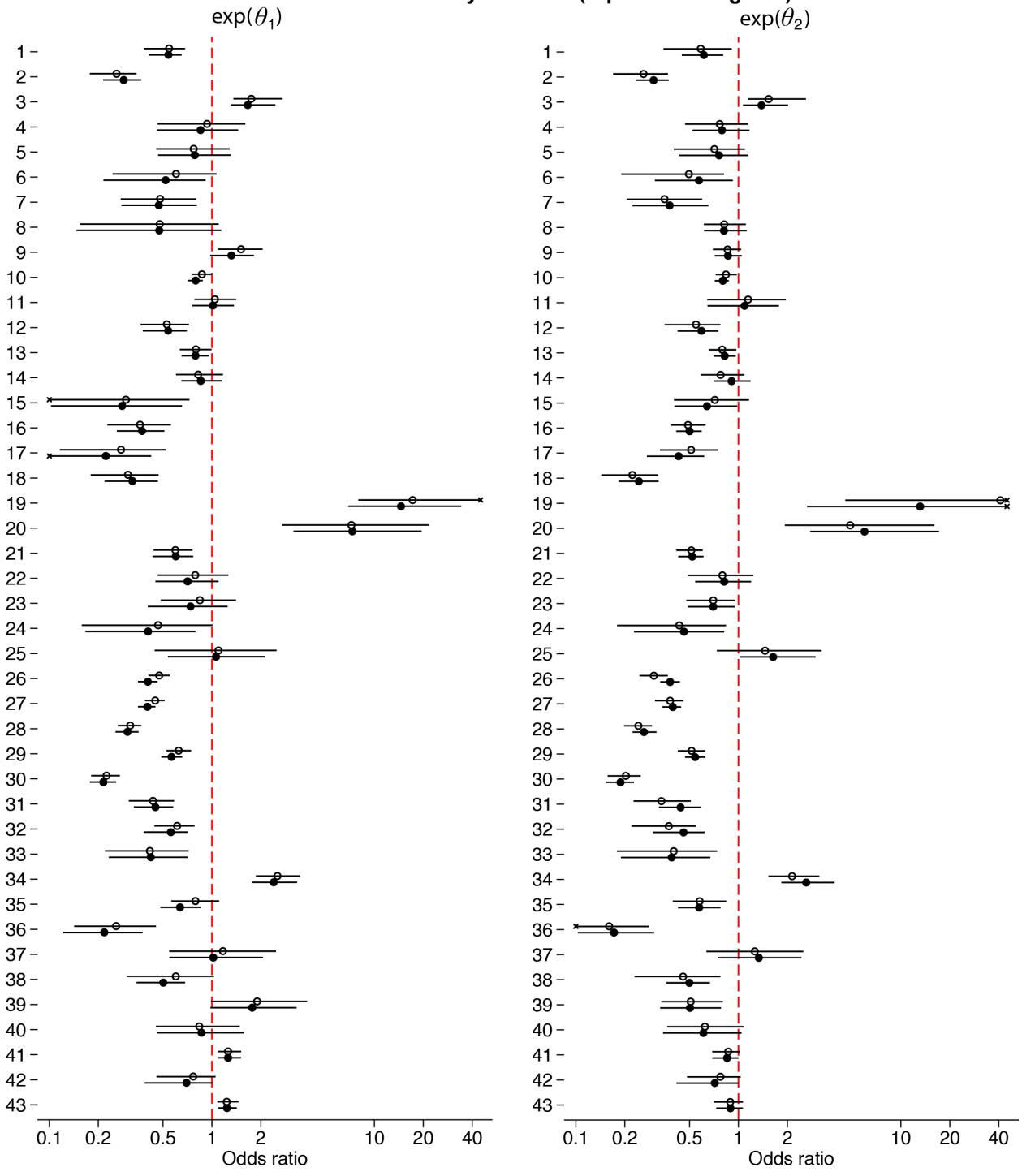
Quantity or topic	Noninformative prior	Prior informative on the sign of the correlation
10	0.40 (-0.85, 0.97)	0.77 (-0.70, 0.99)
11	0.91 (0.27, 0.99)	0.94 (0.54, 0.99)
12	0.87 (-0.11, 0.99)	0.92 (0.09, 0.99)
13	0.52 (-0.81, 0.98)	0.80 (-0.57, 0.99)
14	0.66 (-0.80, 0.98)	0.87 (-0.58, 0.99)
15	0.68 (-0.32, 0.98)	0.84 (-0.02, 0.99)
16	0.90 (0.36, 0.99)	0.93 (0.54, 0.99)
17	0.63 (-0.27, 0.93)	0.74 (-0.03, 0.95)
18	0.24 (-0.91, 0.96)	0.62 (-0.84, 0.98)
19	0.15 (-0.92, 0.95)	0.61 (-0.79, 0.98)
20	0.77 (-0.29, 0.98)	0.88 (0.12, 0.99)
21	-0.20 (-0.95, 0.83)	0.24 (-0.84, 0.96)
22	0.40 (-0.90, 0.97)	0.79 (-0.67, 0.99)
23	0.15 (-0.91, 0.92)	0.49 (-0.78, 0.97)
24	0.48 (-0.90, 0.98)	0.84 (-0.53, 0.99)
25	0.52 (-0.87, 0.97)	0.78 (-0.56, 0.99)
26	0.98 (0.93, 0.99)	0.99 (0.95, 1.00)
27	0.91 (-0.47, 0.99)	0.91 (-0.47, 0.99)
28	0.78 (-0.58, 0.98)	0.89 (-0.02, 0.99)
29	0.43 (-0.92, 0.98)	0.80 (-0.70, 0.99)
30	0.75 (-0.82, 0.98)	0.88 (-0.18, 0.99)
31	0.55 (-0.81, 0.98)	0.85 (-0.48, 0.99)
32	0.27 (-0.89, 0.97)	0.64 (-0.75, 0.98)
33	0.71 (-0.62, 0.98)	0.88 (-0.15, 0.99)
34	-0.01 (-0.94, 0.89)	0.43 (-0.79, 0.99)
35	0.19 (-0.96, 0.97)	0.70 (-0.71, 0.99)
36	0.26 (-0.90, 0.96)	0.68 (-0.72, 0.99)
37	0.29 (-0.90, 0.97)	0.69 (-0.78, 0.99)
38	0.59 (-0.87, 0.98)	0.89 (-0.35, 0.99)
39	-0.80 (-0.98, 0.11)	Not run*
40	0.53 (-0.61, 0.97)	Not run*
41	-0.60 (-0.98, 0.71)	Not run*
42	0.73 (-0.45, 0.98)	Not run*

Table 15. Posterior median and 95% credible interval for between-study correlations with uninformative priors and priors informative on the correlation sign (continued)

Quantity or topic	Noninformative prior	Prior informative on the sign of the correlation
43	-0.20 (-0.96, 0.92)	Not run*
44	0.44 (-1.00, 1.00)	0.81 (0.07, 1.00)
45	0.99 (0.42, 1.00)	0.99 (0.18, 1.00)
ρ_{13}		
44	0.32 (-1.00, 1.00)	0.80 (0.06, 1.00)
45	0.98 (0.17, 1.00)	0.97 (-0.08, 1.00)
ρ_{23}		
44	0.16 (-0.96, 0.99)	0.78 (0.16, 1.00)
45	0.99 (0.72, 1.00)	0.99 (0.68, 1.00)

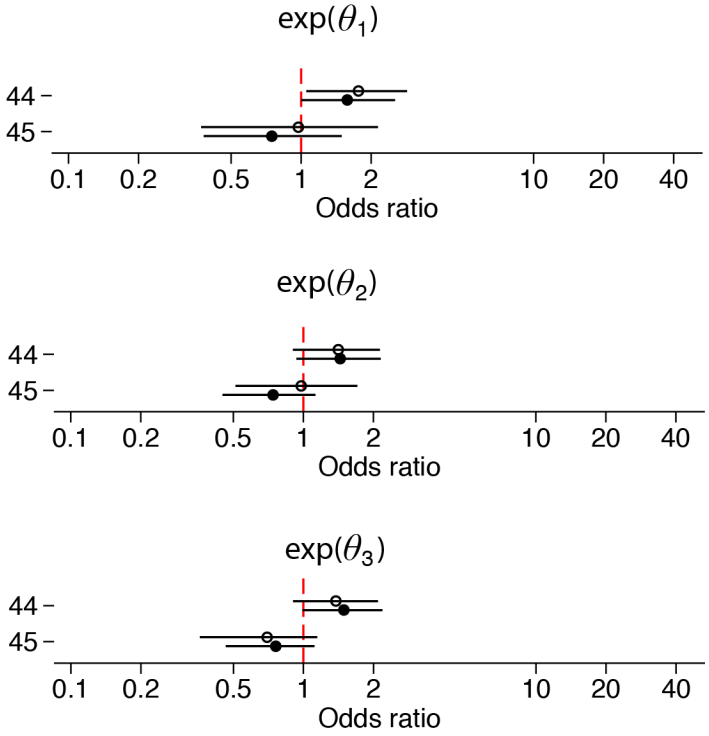
*Topics 39-43 correspond to mutually exclusive outcomes. We have no intuition on whether the between-study correlation among the two log odds ratios is expected to be positive or negative. Recall that although the proportions of people experiencing mutually exclusive outcomes are negatively correlated, the respective treatment effects can be positively or negatively correlated.

Figure 11. Comparison of univariate and multivariate meta-analysis using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43)



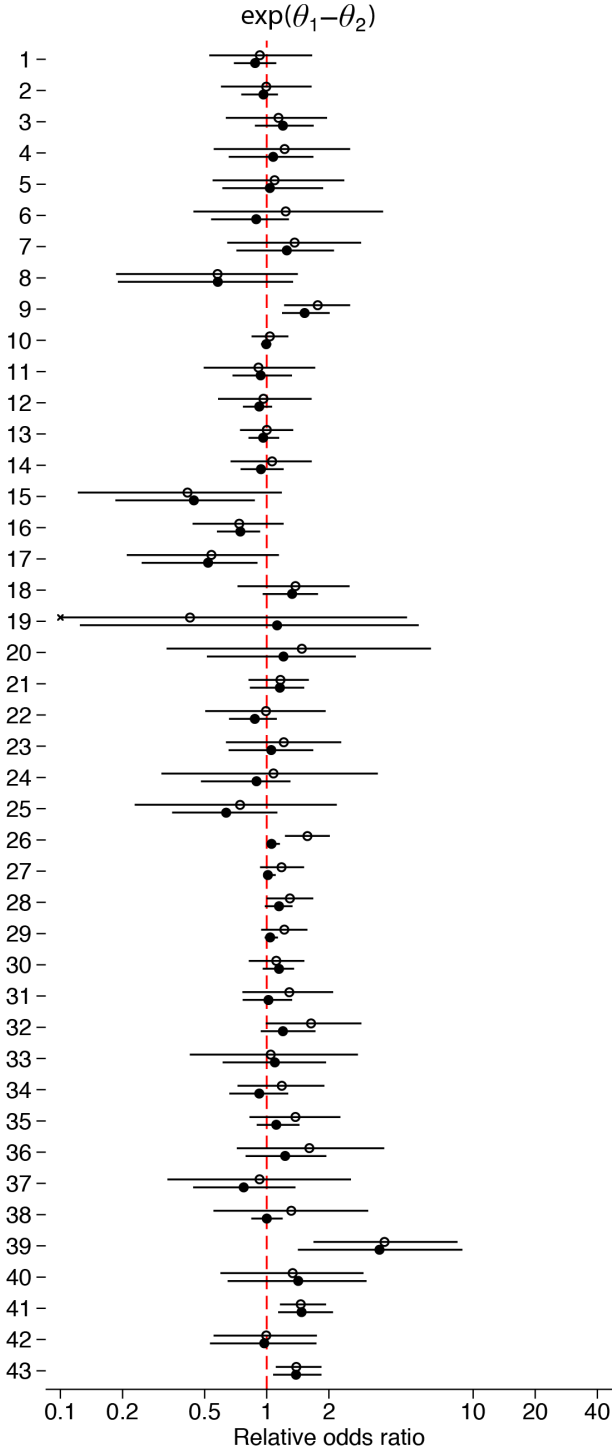
Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 : meta-analysis posterior median for the first outcome; θ_2 : meta-analysis posterior median for the second outcome. For topics 1 through 38 those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive. Small "x" markers denote truncated credible intervals.

Figure 12. Comparison of univariate and trivariate meta-analysis using the binomial or the multinomial distribution to model within-study variance (topics 44 and 45)



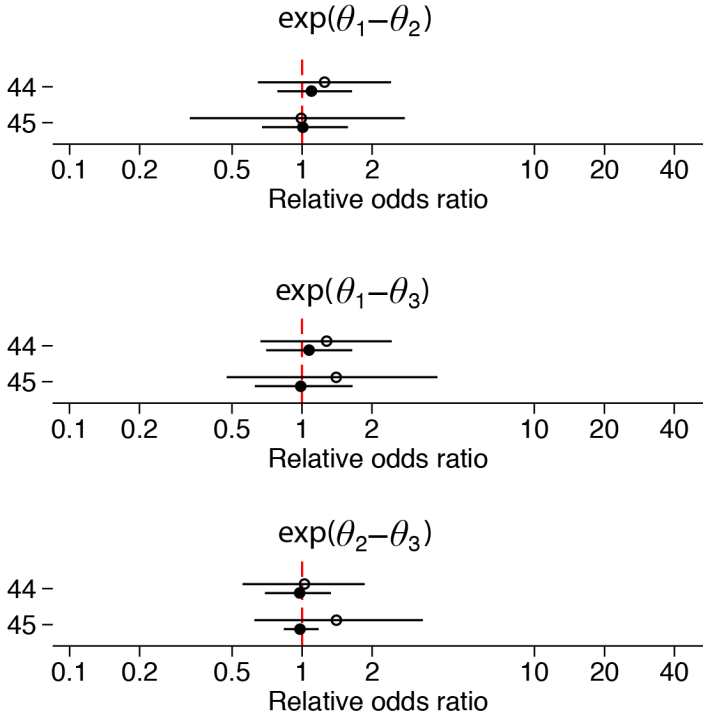
Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 through θ_3 : meta-analysis posterior median for the three outcomes. The three outcomes have an is-subset-of relationship.

Figure 13. Relative summary odds ratios from univariate and bivariate meta-analyses using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43)



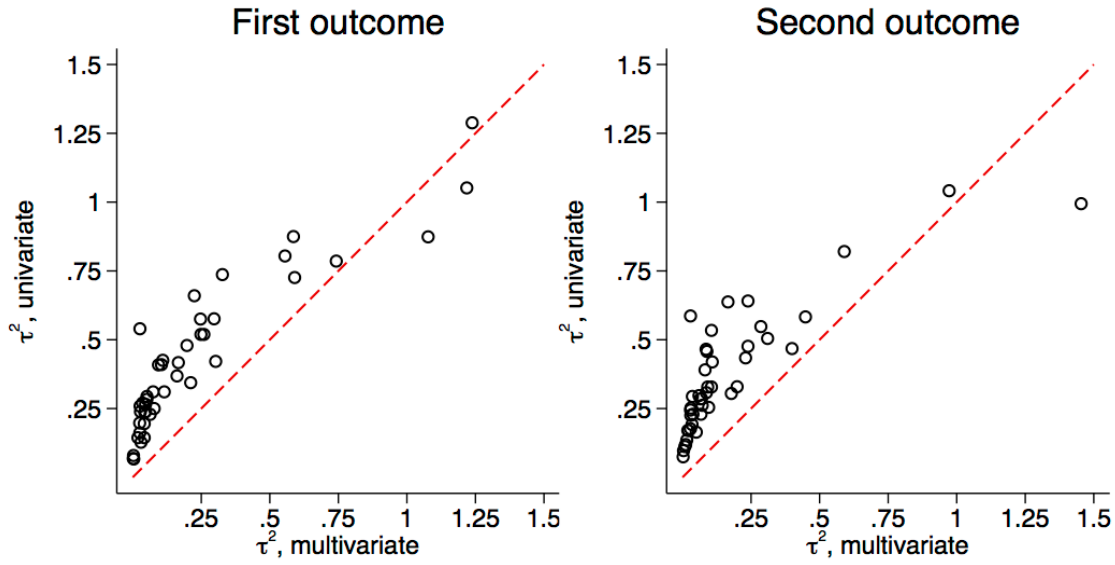
Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 and θ_2: meta-analysis posterior medians for the first and second outcome, respectively. For topics 1 through 38 those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive. Small "x" markers denote truncated credible intervals.

Figure 14. Relative summary odds ratios from univariate and trivariate meta-analyses using the binomial or the multinomial distribution to model within-study variance (topics 44 and 45)



Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 through θ_3: meta-analysis posterior median for the three outcomes. The three outcomes have an is-subset-of relationship.

Figure 15. Between-study variance estimates in univariate versus multivariate analyses in Bayesian analyses



Illustrative Simulation Analyses (Aim 3)

Fidelity of Simulations and Code Integrity

The means and covariances of the simulated counts in the comparator arm and the true effects $\theta_{k,1}$ and $\theta_{k,2}$ matched the simulation parameters closely for all 648 scenarios for mutually exclusive outcomes and all 648 scenarios for outcomes with an is-subset-of relationship. Detailed inspection of graphs of the distribution of the above parameters in scenarios corresponding to combinations of highest and lowest values for parameters, as well as 15 additional randomly chosen scenarios were not suggestive of any systematic errors in programing and analysis.

Exploration of Influential Simulation Parameters With Analysis of Variance (ANOVA)

We used ANOVA to identify influential simulation parameters for the MSE, bias and coverage probability under univariate and multivariate meta-analysis. ANOVA was used as a screening tool. We examined up to two-way interactions between the most influential parameters. Table 16 shows results for simulations of mutually exclusive outcomes. Note that these simulations and their conclusions are symmetric with respect to the two outcomes, in that findings for the second outcome are analogous to findings for the first outcome. Table 17 outlines corresponding explorations for the case of outcomes that have an is-subset-of relationship. Note that the patterns of important factors differ between the first and the second outcome.

Why Are the MSE, Bias and Coverage Dependent on θ_1 or θ_2?

Interestingly, the MSE, bias and coverage depend on θ_1 or θ_2. The reason may not be immediately obvious; if anything, one would expect MSE, bias and coverage to be independent of the true means.

However, recall that in the normal approximation model, the conditional sample covariance matrices (the within-study covariance matrices) are calculated as functions of the proportions of events, and are thus dependent on the true means. Therefore, a correlation between the means and the calculated covariance matrices exists. This situation is analogous to univariate meta-analysis, where for non-variance-stabilizing link functions such as the log() or the logit() functions, the study effects (log risk ratio or log odds ratio, respectively) are correlated with their variances. The multivariate meta-analysis methods that use the multinomial to model within-study variance do not have this shortcoming.

Table 16. Influential simulation parameters for random effects bivariate meta-analysis of two mutually exclusive outcomes

ANOVA factors	SS for $\hat{\theta}_1 \times 10^{-3}$			SS for $\hat{\theta}_2 \times 10^{-3}$		
	MSE	**Coverage**	**Bias**	**MSE**	**Coverage**	**Bias**
K	8.5*	18.7*	–	8.5*	18.5*	–
N	34.1*	1.6*	1.5*	34.1*	2.2*	1.1*
θ_1	0.2*	–	–	–	–	–
θ_2	–	–	–	0.2	–	1.4
$K \times N$	4.0*	–	–	4.0*	–	–
$K \times \theta_1$	6	–	–	–	–	–
$K \times \theta_2$	–	–	–	–	–	–
$N \times \theta_1$	0.1*	–	0.8	–	–	–
$N \times \theta_2$	–	–	–	0.1	–	1.2*
τ_1^2	–	–	–	–	–	–
τ_2^2	–	–	–	–	–	–
ρ	–	–	–	–	–	–

SS: sum of squares of factors in an ANOVA model. The dash stands for factors or interactions with p-values of 0.10 or larger. The asterisks denote factors or interactions with p-value less than 0.05.

Table 17. Influential simulation parameters for random effects bivariate meta-analysis of two outcomes with an is-subset-of relationship

ANOVA factors	SS for $\hat{\theta}_1 \times 10^{-3}$			SS for $\hat{\theta}_2 \times 10^{-3}$		
	MSE	**Coverage**	**Bias**	**MSE**	**Coverage**	**Bias**
K	8.2*	16.2*	–	3.6*	21.1*	–
N	32.7*	1.2	6.3*	15.4*	–	–
θ_1	0.2	–	9.6*	–	–	–
θ_2	–	–	–	–	–	–
$K \times N$	3.8*	–	–	1.8*	–	–
$K \times \theta_1$	–	–	–	–	–	–
$K \times \theta_2$	–	–	–	–	–	–
$N \times \theta_1$	0.1*	–	4.9*	–	–	–
$N \times \theta_2$	–	–	–	0.07	–	–
τ_1^2	–	–	–	–	–	–
τ_2^2	–	–	–	–	–	–
ρ	–	–	–	–	–	–

SS: sum of squares of factors in an ANOVA model. The dash stands for factors or interactions with p-values of 0.10 or larger. The asterisks denote factors or interactions with p-value less than 0.05.

MSE, Bias and Coverage for Bivariate Random Effects Meta-Analysis

Based on the explorative analyses in Table 16 and Table 17 we report the MSE, bias and coverage averaging over choices for between-study heterogeneity and between-study correlations in the true means. Table 18 shows these results for two mutually exclusive outcomes.

Table 19 shows the results for two outcomes having an is-subset-of relationship. Overall, the pattern of results is similar in the two tables. Note that the coverage is consistently above the desired 95 percent, which suggests that the t distribution with one degree of freedom is somewhat conservative for the scenarios examined here. Results for scenarios where half the studies reported both outcomes and half reported only the first outcomes were similar (not shown).

Table 20 and Table 21 are the corresponding results for τ_1^2, τ_2^2 and ρ averaging over the factors that were most influential in the ANOVA exploratory analyses (the number of studies K, the sample size per study N, and the true means in each outcome θ_1 and θ_2). Note that the MSE, bias and coverage do not vary very much within each column, as expected. The Appendix has the corresponding results for univariate meta-analyses.

Table 18. MSE, coverage and bias for random effects bivariate meta-analysis (two mutually exclusive outcomes) averaging over τ_1^2, τ_2^2 and ρ.

K	N	$\exp(\theta_1)$	$\exp(\theta_2)$	MSE for $\hat{\theta}_1$ ($\times 10^{-3}$)	MSE for $\hat{\theta}_2$ ($\times 10^{-3}$)	Bias for $\hat{\theta}_1$ ($\times 10^{-3}$)	Bias for $\hat{\theta}_2$ ($\times 10^{-3}$)	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
10	50	1	1	28	28	-2	0	98.8	98.8
10	50	1	1.5	28	25	0	-6	98.7	98.8
10	50	1.5	1	25	28	-5	-1	98.7	98.9
10	50	1.5	1.5	25	25	-4	-5	98.7	98.8
10	100	1	1	14	14	0	-1	98.6	98.9
10	100	1	1.5	14	13	0	-1	98.8	98.6
10	100	1.5	1	13	14	-2	-1	98.6	98.5
10	100	1.5	1.5	13	13	-1	-2	98.5	98.4
10	500	1	1	3	3	0	0	98.4	98.3
10	500	1	1.5	3	3	0	-1	98.6	98.5
10	500	1.5	1	3	3	0	1	98.4	98.5
10	500	1.5	1.5	3	3	0	0	98.5	98.3
20	50	1	1	14	14	0	1	98.1	97.9
20	50	1	1.5	14	13	-1	-9	98.0	97.7
20	50	1.5	1	13	13	-11	1	97.5	98.0
20	50	1.5	1.5	12	12	-7	-6	97.8	97.6
20	100	1	1	7	7	0	-1	97.5	97.6
20	100	1	1.5	7	6	1	-3	97.4	97.5
20	100	1.5	1	6	7	-3	2	97.4	97.4
20	100	1.5	1.5	6	6	-2	-2	97.5	97.4
20	500	1	1	1	1	0	0	97.4	97.4
20	500	1	1.5	1	1	1	-1	97.2	97.4
20	500	1.5	1	1	1	0	0	97.5	97.4
20	500	1.5	1.5	1	1	-1	0	97.3	97.1

Table 19. MSE, coverage and bias for random effects bivariate meta-analysis (two outcomes that have an is-subset-of relationship) averaging over τ_1^2, τ_2^2 and ρ

K	N	$\exp(\theta_1)$	$\exp(\theta_2)$	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
10	50	1	1	27	18	-2	-1	98.9	98.7
10	50	1	1.5	28	17	-1	1	98.7	98.6
10	50	1.5	1	25	18	-15	2	98.7	98.5
10	50	1.5	1.5	25	17	-11	2	98.7	98.6
10	100	1	1	14	9	0	-1	98.7	98.5
10	100	1	1.5	14	9	0	2	98.8	98.5
10	100	1.5	1	13	9	-7	-1	98.5	98.6
10	100	1.5	1.5	13	9	-5	1	98.6	98.5
10	500	1	1	3	2	0	0	98.4	98.5
10	500	1	1.5	3	2	0	0	98.5	98.5
10	500	1.5	1	3	2	-2	0	98.5	98.5
10	500	1.5	1.5	3	2	-1	0	98.4	98.5
20	50	1	1	13	9	0	1	98.0	97.5
20	50	1	1.5	13	9	-2	-1	98.1	97.3
20	50	1.5	1	12	9	-23	0	97.3	97.4
20	50	1.5	1.5	12	9	-15	2	97.8	97.3
20	100	1	1	7	4	-1	-1	97.9	97.6
20	100	1	1.5	7	4	1	1	97.7	97.4
20	100	1.5	1	6	4	-10	1	97.4	97.4
20	100	1.5	1.5	6	4	-6	0	97.6	97.4
20	500	1	1	1	1	0	0	97.3	97.5
20	500	1	1.5	1	1	0	0	97.3	97.3
20	500	1.5	1	1	1	-2	0	97.5	97.2
20	500	1.5	1.5	1	1	-2	0	97.6	97.5

Table 20. MSE, coverage and bias for random effects bivariate meta-analysis (two mutually exclusive outcomes) averaging over K, N, θ_1 and θ_2

τ_1^2	τ_2^2	ρ	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
0	0	-0.8	10.5	10.2	-1.9	0.0	98.2	98.0
0	0	-0.5	10.3	10.8	-2.3	-2.0	98.1	98.1
0	0	-0.2	10.8	10.6	-1.0	-2.2	98.1	98.0
0	0.1	-0.8	10.5	10.8	-1.9	-2.2	98.3	97.9
0	0.1	-0.5	10.8	10.4	-0.7	-0.6	98.0	98.3
0	0.1	-0.2	10.5	10.7	-1.9	-3.0	98.0	98.0
0	0.5	-0.8	10.7	10.8	-0.4	-0.4	98.0	98.1
0	0.5	-0.5	10.6	10.7	-1.4	-2.0	98.1	98.1
0	0.5	-0.2	10.7	10.6	-1.9	-1.5	98.1	98.2
0.1	0	-0.8	10.7	10.8	-2.4	-1.5	98.3	98.0
0.1	0	-0.5	10.6	10.7	-2.9	-0.4	98.2	98.1
0.1	0	-0.2	10.6	10.6	-1.1	-2.5	98.0	97.9
0.1	0.1	-0.8	10.8	10.4	-2.1	-1.1	97.9	98.0
0.1	0.1	-0.5	10.5	10.9	-2.8	-1.4	98.1	97.8
0.1	0.1	-0.2	10.6	10.7	0.0	-1.7	98.2	98.0
0.1	0.5	-0.8	10.9	11.0	-1.7	-3.2	97.9	97.9
0.1	0.5	-0.5	10.6	10.7	-0.3	-3.1	98.0	98.0
0.1	0.5	-0.2	10.8	11.0	-2.0	-0.4	98.0	98.0
0.5	0	-0.8	10.6	10.6	-3.3	-0.8	97.9	98.1
0.5	0	-0.5	10.5	10.9	-0.4	-1.5	98.1	98.0
0.5	0	-0.2	10.5	10.5	-0.7	-1.7	97.9	98.2
0.5	0.1	-0.8	10.7	10.4	0.7	-2.2	98.0	98.2
0.5	0.1	-0.5	11.0	10.6	-1.7	-1.0	98.1	98.3
0.5	0.1	-0.2	10.8	10.4	-1.8	-1.5	98.0	98.3
0.5	0.5	-0.8	10.9	10.7	-2.2	-0.6	98.0	98.0
0.5	0.5	-0.5	10.6	10.6	-1.5	0.3	98.2	98.1
0.5	0.5	-0.2	10.8	10.3	-3.0	-0.7	98.1	98.3

Table 21. MSE, coverage and bias for random effects bivariate meta-analysis (two outcomes that have an is-subset-of relationship) averaging over K, N, θ_1 and θ_2

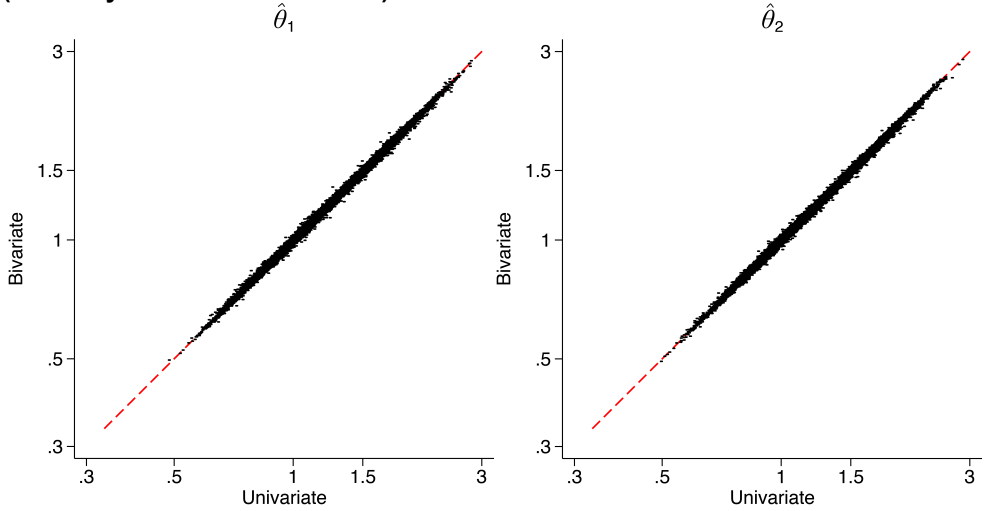
τ_1^2	τ_2^2	ρ	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
0	0	-0.8	10.5	7.1	-4.8	-1.2	98.0	98.1
0	0	-0.5	10.1	6.9	-4.1	0.9	98.2	98.0
0	0	-0.2	10.4	7.1	-5.6	1.2	98.3	97.9
0	0.1	-0.8	10.2	7.1	-4.0	-0.2	97.9	98.0
0	0.1	-0.5	10.8	7.1	-3.0	1.7	98.0	97.9
0	0.1	-0.2	10.4	7.0	-3.7	0.8	98.1	97.9
0	0.5	-0.8	10.7	7.0	-5.3	-1.6	98.1	97.7
0	0.5	-0.5	10.2	6.9	-3.5	0.7	98.2	97.9
0	0.5	-0.2	10.5	7.2	-1.9	2.7	98.0	97.9
0.1	0	-0.8	10.5	6.9	-4.9	-0.2	98.0	98.0
0.1	0	-0.5	10.4	7.0	-4.3	0.7	97.9	97.8
0.1	0	-0.2	10.5	7.0	-4.1	0.7	98.2	98.1
0.1	0.1	-0.8	10.4	7.2	-3.0	1.7	98.1	97.9
0.1	0.1	-0.5	10.4	7.2	-6.5	-0.4	98.1	97.6
0.1	0.1	-0.2	10.5	6.9	-4.5	-0.5	98.0	98.1
0.1	0.5	-0.8	10.6	7.0	-5.5	-0.1	98.0	98.0
0.1	0.5	-0.5	10.3	7.0	-2.4	0.9	98.2	98.0
0.1	0.5	-0.2	10.7	7.2	-6.0	-0.7	98.2	97.9
0.5	0	-0.8	10.3	6.9	-4.3	0.5	98.1	98.0
0.5	0	-0.5	10.5	6.9	-4.4	0.5	98.3	98.2
0.5	0	-0.2	10.5	7.2	-6.5	-1.5	98.0	97.9
0.5	0.1	-0.8	10.5	7.0	-4.5	-0.1	98.0	98.0
0.5	0.1	-0.5	10.7	7.1	-3.9	1.6	98.3	98.2
0.5	0.1	-0.2	10.5	7.1	-3.4	0.7	98.0	97.8
0.5	0.5	-0.8	10.5	7.2	-4.6	0.4	98.3	98.1
0.5	0.5	-0.5	10.5	7.0	-5.5	0.2	98.2	97.9
0.5	0.5	-0.2	10.5	7.2	-4.4	0.5	98.1	98.0

Comparison Between Univariate and Multivariate Meta-Analysis in Simulations

Comparison of Point Estimates

The point estimates for summary log odds ratios were very similar under univariate and multivariate meta-analysis. This is true for mutually exclusive outcomes (Figure 16), for outcomes with an is-subset-of relationship (Figure 17), and for sensitivity analysis scenarios where half of the studies did not report results for the second outcome (not shown). This finding is congruent with the conclusions of the empirical analyses.

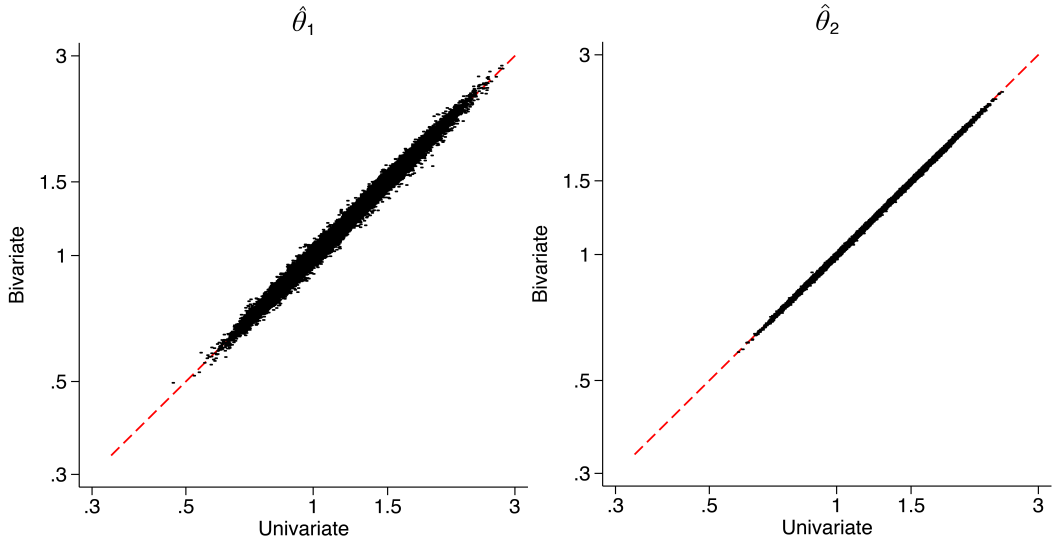
Figure 16. Summary odds ratios from univariate and bivariate meta-analysis across all simulations (mutually exclusive outcomes)



The axes correspond to odds ratios. The reference line is the line of equality.

In Figure 16 the scatter of the simulation points is identical for the two outcomes. This is because for mutually exclusive outcomes simulation results and conclusions are exactly symmetric with respect to the two outcomes. By contrast, for outcomes with an is-subset-of relationship (Figure 17), the random scatter of the points is greater for the first outcome than for the second outcome. This is expected, as the number of those experiencing the second outcome within each simulated study arm is larger, resulting in lower sampling variance.

Figure 17. Summary odds ratios from univariate and bivariate meta-analysis across all simulations (outcomes with an is-subset-of relationship)

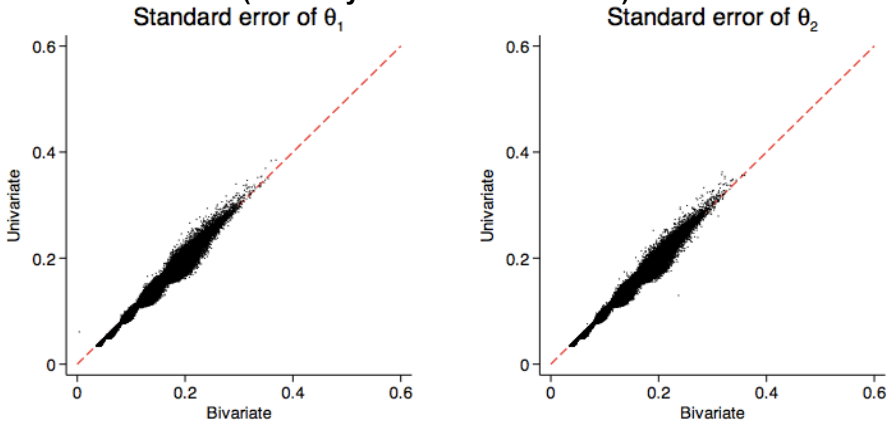


The axes correspond to odds ratios. The reference line is the line of equality.

Comparison of Standard Errors for the Mean Effects

As shown in Figure 18 for mutually exclusive outcomes and Figure 19 for outcomes with an is-subset-of relationship, the standard errors of the meta-analysis means are not substantially different in univariate versus bivariate analyses. This is congruent with the empirical results.

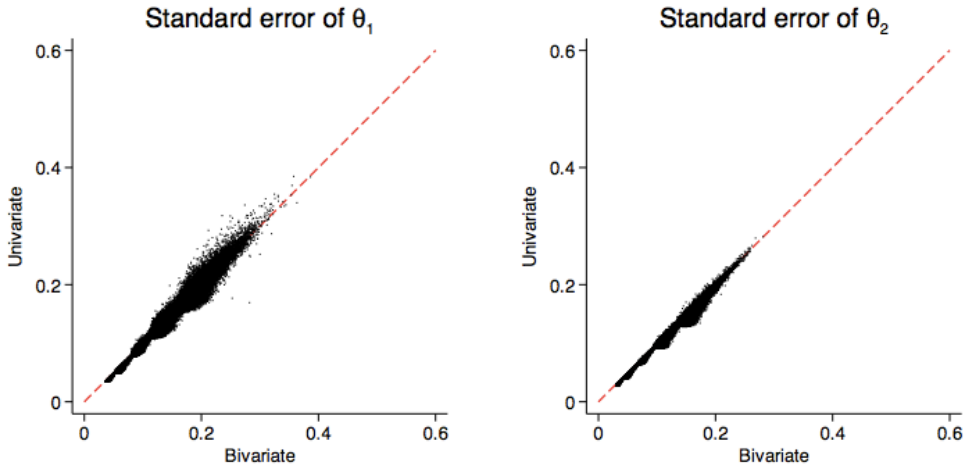
Figure 18. Standard errors of summary odds ratios with univariate and bivariate meta-analysis across all simulations (mutually exclusive outcomes)



The reference line is the line of equality.

Again, note that the two panels in Figure 18 are identical, because the simulation is symmetric with respect to the two mutually exclusive outcomes. In Figure 19, the standard errors for the second outcome are smaller than those for the first outcome; the explanation is the same as for Figure 17.

Figure 19. Standard errors of summary odds ratios from univariate and bivariate meta-analysis across all simulations (outcomes with an is-subset-of relationship)

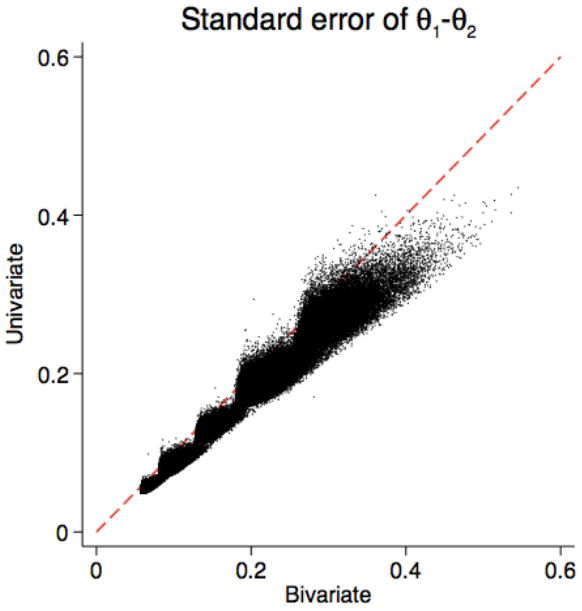


The reference line is the line of equality.

Comparison of Standard Errors of Differences in the Mean Effects

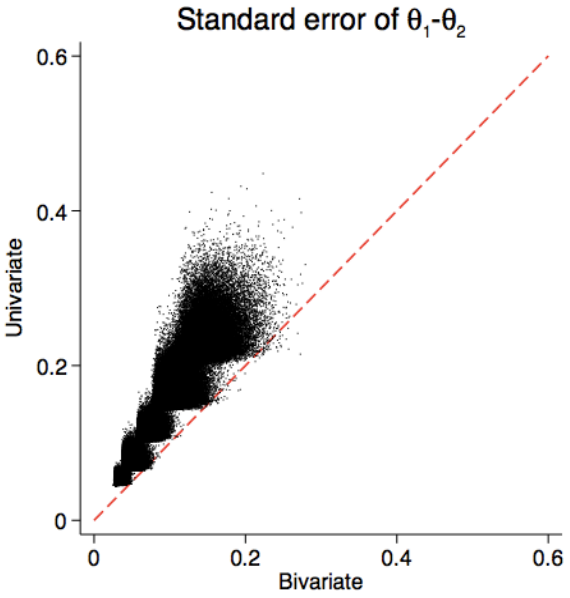
The most pronounced differences between univariate and bivariate meta-analyses pertain to the standard errors of differences in the summary estimates (or standard errors of linear combinations of the summary estimates in general[5,11]). In simulations of mutually exclusive outcomes, the between-study correlation is most often negative, and thus the standard error of the difference in the two means tends to be larger with bivariate compared with univariate meta-analysis (Figure 20). In simulations of outcomes that have an is-subset-of relationship the standard error with univariate meta-analyses tends to be larger than the standard error with multivariate meta-analyses (Figure 21), because the correlations between the outcomes were simulated to be positive, and are estimated as such. The patterns evident in the scatterplots correspond to combinations of simulation parameters, mainly study sample size *N* and the number of studies in a meta-analysis *K*.

Figure 20. Comparison of standard errors of the difference in the log summary odds ratios of two mutually exclusive outcomes with univariate and bivariate meta-analysis



The reference line is the line of equality.

Figure 21. Comparison of standard errors of the difference in the log summary odds ratios of two outcomes that have an is-subset-of relationship with univariate and bivariate meta-analysis



The reference line is the line of equality.

Discussion

We performed a large-scale empirical comparison of univariate and multivariate meta-analysis using data from the Cochrane Library of Systematic Reviews, and complemented it with a simulation study. Overall, univariate and multivariate methods yield numerically similar means and confidence intervals, suggesting that systematic review conclusions are not sensitive to this particular choice of methods. However, the confidence intervals of relative odds ratios between the pairs of outcomes can differ substantially between univariate and multivariate meta-analysis.

It appears that, as long as we focus on summaries for individual outcomes (and the respective confidence intervals) the choice between univariate and multivariate meta-analysis has limited practical importance. This is supported by our simulation analyses, and is congruent with the numerical results in the worked examples of several methodological papers introducing or reviewing methods for multivariate meta-analysis.[2,4,5,7,8,11,13,48,49] It is not clear whether this would be observed in other examples or in other types of data, where the information on within-study correlations of treatment effects is not extractable (but presumably available from external sources, e.g., by contacting authors). Because the actual mechanics of meta-analysis methodologies are the same, however, it is likely that similar observations would hold for a wider range of examples.

It is not clear how often systematic reviewers face the methodological dilemma explored in this work. In our empirical evaluation, out of 1919 reviews with at least one binary meta-analysis, 29 (1.5 percent) reviews had at least one pair of meta-analyses that fulfilled our eligibility criteria. This proportion is probably an underestimate. We used outcomes exactly as defined by the Cochrane reviewers, and did not make any effort to redefine them to facilitate joint meta-analysis. Cochrane reviews include only univariate meta-analysis; if they were routinely performing multivariate analyses, they might have reviewed a larger number of outcomes. Further, we limited our analyses to examples where counts of combinations of outcomes are exactly recoverable from data used in univariate meta-analysis. However, it is possible that complete data on combinations of categorical outcomes can be obtained by contacting primary study authors, or even with care and perseverance during extraction of data from published articles.[50] Finally, a single reviewer judged the eligibility of each pair or triplet of meta-analyses from the Cochrane Library, without checking by others. Nevertheless, eligibility criteria pertaining to the number of total studies or studies that are common to all outcomes, and the minimum number of patients were done programmatically, and thus consistently. The only judgment calls pertained to the relationships between pairs or triplets of outcomes (mutually exclusive, one being a subset of the other, or other relationship).

Our results and conclusions are limited by the decision to use multivariate outcomes that could be represented as a set of categories, either mutually exclusive or represented as subsets of each other. This choice was motivated by the desire to have known correlations among the multivariate outcomes, but it does rule out consideration of many common multivariate outcomes and design structures for which our findings may not hold. Common multivariate outcomes that we do not consider include different biomarkers, repeated measurements of outcomes at different times, different adverse events, combinations of efficacy and safety endpoints, combinations of medical outcomes and quality of life measures and bivariate analysis of sensitivity and specificity in studies of diagnostic test accuracy.

Another key observation is that although the Cochrane reviews reported outcomes as event counts either at one or several points in time, the outcomes are fundamentally not counts but rather time-to-event outcomes that could, or perhaps should, be analyzed by survival analysis

with appropriate adjustment for censoring. Moreover, the different types of outcome categories suggest competing risks analysis. While such analyses may be preferred if the individual outcome times are available, in many cases only summary counts are reported and time-to-event analyses must be sacrificed in place of the multinomial analyses used here.

Thus, our conclusions must be tempered by the restricted set of problems considered, the lack of reporting of appropriate metrics of analysis, the lack of complete individual patient data with which to carry out the ideal statistical analysis and the necessarily limited simulations that, for instance, only consider two outcomes at a time. Further study may uncover differences between the univariate and multivariate analyses that we did not find. Of particular importance, broader conclusions can be drawn through analytical approaches, at least in the models that use the normal approximation to model within study variance.

If the patterns that we observed in this work are more broadly applicable, and provided that one is not interested in linear combinations of treatment effects across outcomes (e.g., log relative odds ratios), it may be argued that decisions between univariate (separate) and multivariate (joint) meta-analysis have theoretical rather than practical interest. So should one use separate or joint meta-analysis for sets of outcomes that can be approached with either method? In theory, the decision on performing separate versus joint meta-analyses depends on the underlying assumptions that the researcher is prepared to make about the data. Ideally, these decisions should be made early in the analysis, and not after examining the data. The key reason for using multivariate meta-analysis is that, through the correlations, it utilizes more information. Though in the majority of the 45 applications there is very little clinical or statistical difference in the results/conclusions, this itself is an important finding in each case. In any single application, if the multivariate approach does not change the conclusions from a univariate approach this increases the reliability of the findings, and gives some reassurance to the clinician that the findings robust. The fact that the conclusion does not change does not automatically render the multivariate result of no practical use (see also the discussion by Trikalinos and Olkin[11]).

An additional opportunity where multivariate meta-analysis may yield more precise or different results than univariate analyses, is when there is preferential non-reporting of results for one of the outcomes that could be analyzed jointly. Many systematic reviews neglect to analyze certain outcomes because of the number of studies in which these outcomes go unreported. The remaining studies may be felt either to be too few to provide an accurate estimate or to be unrepresentative of the complete set because of outcome reporting bias caused by failure to report the outcome because of the lack of statistical or biological significance of its estimated effect. Because multivariate models incorporate the correlations between the outcomes, they may provide information about the missing outcomes and enable them to be effectively incorporated into analyses by the borrowing of strength from the observed outcomes.[51] In such cases, multivariate models may give more accurate and more precise estimates than univariate models.

While we consider it commendable to conduct both univariate and multivariate meta-analysis in sensitivity analyses, when possible, we are reluctant to recommend this practice as a minimum standard for systematic review and meta-analysis. A minimum standard implies that failing to follow the recommendation can result in misleading conclusions, and prove detrimental to decision making. In our opinion, our findings and the findings of others are compatible with the notion that using multivariate meta-analysis methods is good practice, but probably not a prerequisite for drawing valid conclusions in an applied meta-analysis setting.

References

1. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Stat Med. Jan 26 2011. [epub ahead of print]
2. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, eds. The handbook of research synthesis. Vol 2nd. New York: Russel Sage Foundation; 1994:339-55.
3. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Stat Med. Dec 30 1993;12(24):2273-84.
4. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002;21(4):589-624.
5. Trikalinos TA, Olkin I. A method for the meta-analysis of mutually exclusive binary outcomes. Stat Med. 2008;27(21):4279-300.
6. Arends LR. Multivariate meta-analysis: Modeling the heterogeneity (PhD Thesis). Haveka BV: Alblasserdam. 2006.
7. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. Stat Med. 1996;15(5):537-57.
8. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, et al. Meta-analysis of multiple outcomes by regression with random effects. Stat Med. 1998;17(22):2537-50.
9. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. Stat Med. May 30 2010;29(12):1282-97.
10. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. Stat Methods Med Res. Feb 16 2012. [Epub ahead of print].
11. Trikalinos TA, Olkin I. Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. Clin Trials. 2012;9(5):610-20.
12. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation J Royal Stat Soc A. 2009;172(4):789-811.
13. Peter I, Crosier MD, Yoshida M, et al. Associations of APOE gene polymorphisms with bone mineral density and fracture risk: a meta-analysis. Osteoporosis Int. Apr 2011;22(4):1199-209.
14. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. J Clin Epidemiol. Jan 2008;61(1):41-51.
15. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. Stat Med. Dec 20 2010;29(29):3046-3067.
16. White IR. Multivariate random effects meta-regression. Stata J. 2011;11:255-70.
17. Greenhalgh J, Hockenhull J, Rao N, et al. Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes. Cochrane Database Syst Rev. 2010(5):CD004587.
18. Rostom A, Dube C, Wells G, et al. Prevention of NSAID-induced gastroduodenal ulcers. Cochrane Database Syst Rev. 2002(4):CD002296.
19. Halliday HL, Ehrenkranz RA, Doyle LW. Late (>7 days) postnatal corticosteroids for chronic lung disease in preterm infants. Cochrane Database Syst Rev. 2009(1):CD001145.
20. Barrington KJ, Finer N. Inhaled nitric oxide for respiratory failure in preterm infants. Cochrane Database Syst Rev. 2010(12):CD000509.

21. Gong Y, Huang ZB, Christensen E, et al. Ursodeoxycholic acid for primary biliary cirrhosis. Cochrane Database Syst Rev. 2008(3):CD000551.

22. Hodson EM, Willis NS, Craig JC. Corticosteroid therapy for nephrotic syndrome in children. Cochrane Database Syst Rev. 2007(4):CD001533.

23. Brion LP, Bell EF, Raghuveer TS. Vitamin E supplementation for prevention of morbidity and mortality in preterm infants. Cochrane Database Syst Rev. 2003(4):CD003665.

24. Brown J, PM OB, Marjoribanks J, et al. Selective serotonin reuptake inhibitors for premenstrual syndrome. Cochrane Database Syst Rev. 2009(2):CD001396.

25. Al-Inany HG, Youssef MA, Aboulghar M, et al. Gonadotrophin-releasing hormone antagonists for assisted reproductive technology. Cochrane Database Syst Rev. 2011(5):CD001750.

26. Draper BH, Morroni C, Hoffman M, et al. Depot medroxyprogesterone versus norethisterone oenanthate for long-acting progestogenic contraception. Cochrane Database Syst Rev. 2006(3):CD005214.

27. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database Syst Rev. 2006(3):CD004454.

28. Renfrew MJ, McCormick FM, Wade A, et al. Support for healthy breastfeeding mothers with healthy term babies. Cochrane Database Syst Rev. 2012;5:CD001141.

29. Hofmeyr GJ, Xu H. Amnioinfusion for meconium-stained liquor in labour. Cochrane Database Syst Rev. 2010(1):CD000014.

30. Soares-Weiser K, Goldberg E, Tamimi G, et al. Rotavirus vaccine for preventing diarrhoea. Cochrane Database Syst Rev. 2004(1):CD002848.

31. Hodson EM, Craig JC, Strippoli GF, et al. Antiviral medications for preventing cytomegalovirus disease in solid organ transplant recipients. Cochrane Database Syst Rev. 2008(2):CD003774.

32. Wilhelmus KR. Antiviral treatment and other therapeutic interventions for herpes simplex virus epithelial keratitis. Cochrane Database Syst Rev. 2010(12):CD002898.

33. Webster AC, Ruster LP, McGee R, et al. Interleukin 2 receptor antagonists for kidney transplant recipients. Cochrane Database Syst Rev. 2010(1):CD003897.

34. Webster A, Woodroffe RC, Taylor RS, et al. Tacrolimus versus cyclosporin as primary immunosuppression for kidney transplant recipients. Cochrane Database Syst Rev. 2005(4):CD003961.

35. Schieppati A, Perna A, Zamora J, et al. Immunosuppressive treatment for idiopathic membranous nephropathy in adults with nephrotic syndrome. Cochrane Database Syst Rev. 2004(4):CD004293.

36. Furukawa T, McGuire H, Barbui C. Low dosage tricyclic antidepressants for depression. Cochrane Database Syst Rev. 2003(3):CD003197.

37. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. Cochrane Database Syst Rev. 2006(3):CD004125.

38. Bell RF, Dahl JB, Moore RA, et al. Perioperative ketamine for acute postoperative pain. Cochrane Database Syst Rev. 2006(1):CD004603.

39. Stacey D, Bennett CL, Barry MJ, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2011(10):CD001431.

40. Oates-Whitehead RM, Baumer JH, Haines L, et al. Intravenous immunoglobulin for the treatment of Kawasaki disease in children. Cochrane Database Syst Rev. 2003(4):CD004000.

41. Hodnett ED, Gates S, Hofmeyr GJ, et al. Continuous support for women during childbirth. Cochrane Database Syst Rev. 2011(2):CD003766.

42. Duley L, Henderson-Smart DJ, Meher S, et al. Antiplatelet agents for preventing pre-eclampsia and its complications. Cochrane Database Syst Rev. 2007(2):CD004659.

43. Hatem M, Sandall J, Devane D, et al. Midwife-led versus other models of care for childbearing women. Cochrane Database Syst Rev. 2008(4):CD004667.

44. Furukawa TA, Streiner DL, Young LT. Antidepressant and benzodiazepine for major depression. Cochrane Database Syst Rev. 2002(1):CD001026.

45. Amsallem E, Kasparian C, Haddour G, et al. Phosphodiesterase III inhibitors for heart failure. Cochrane Database Syst Rev. 2005(1):CD002230.

46. Valgimigli M, Campo G, Arcozzi C, et al. Two-year clinical follow-up after sirolimus-eluting versus bare-metal stent implantation assisted by systematic glycoprotein IIb/IIIa Inhibitor Infusion in patients with myocardial infarction: results from the STRATEGY study. J Am Coll Cardiol. Jul 10 2007;50(2):138-45.

47. Valgimigli M, Percoco G, Malagutti P, et al. Tirofiban and sirolimus-eluting stent vs abciximab and bare-metal stent for acute myocardial infarction: a randomized trial. JAMA. May 4 2005;293(17):2109-17.

48. Riley RD, Abrams KR, Lambert PC, et al. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat Med. Jan 15 2007;26(1):78-97.

49. Riley RD, Abrams KR, Sutton AJ, et al. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC Med Res.Methodol. 2007;7:3.

50. Trikalinos TA, Hoaglin DC, Small KM, et al. Methods for the joint meta-analysis of multiple tests. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

51. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Stat Med. Sep 10 2012;31(20):2179-95.

Appendix A. Formulas, Figures, and Tables

Formulas for variances and covariances of log odds ratios for mutually exclusive outcomes in study k

Dropping the study index we write $p_1 = \hat{\pi}_1$ and $p_2 = \hat{\pi}_2$ for the estimates of the proportions for the two mutually exclusive outcomes. Then:

$$\hat{\theta}_1 = \log\left(\frac{p_1^{(T)}}{1-p_1^{(T)}}\right) - \log\left(\frac{p_1^{(C)}}{1-p_1^{(C)}}\right),$$

$$\hat{\theta}_2 = \log\left(\frac{p_2^{(T)}}{1-p_2^{(T)}}\right) - \log\left(\frac{p_2^{(C)}}{1-p_2^{(C)}}\right),$$

$$\hat{\text{var}}(\hat{\theta}_1) = \frac{1}{N^{(T)}p_1^{(T)}(1-p_1^{(T)})} + \frac{1}{N^{(C)}p_1^{(C)}(1-p_1^{(C)})},$$

$$\hat{\text{var}}(\hat{\theta}_2) = \frac{1}{N^{(T)}p_2^{(T)}(1-p_2^{(T)})} + \frac{1}{N^{(C)}p_2^{(C)}(1-p_2^{(C)})}, \text{ and}$$

$$\hat{\text{cov}}(\hat{\theta}_1, \hat{\theta}_2) = -\frac{1}{N^{(T)}(1-p_1^{(T)})(1-p_2^{(T)})} - \frac{1}{N^{(C)}(1-p_1^{(C)})(1-p_2^{(C)})}.$$

Formulas for variances and covariances of log odds ratios for outcomes that have an is-subset-of relationship in study k

Dropping the study index we write $l_1 = \hat{\lambda}_1$ and $l_2 = \hat{\lambda}_2$. The number of those experiencing the first outcome (indexed by 1) is a subset of those experiencing the second outcome (indexed by 2). Then:

$$\hat{\theta}_1 = \log\left(\frac{l_1^{(T)}}{1-l_1^{(T)}}\right) - \log\left(\frac{l_1^{(C)}}{1-l_1^{(C)}}\right),$$

$$\hat{\theta}_2 = \log\left(\frac{l_2^{(T)}}{1-l_2^{(T)}}\right) - \log\left(\frac{l_2^{(C)}}{1-l_2^{(C)}}\right),$$

$$\hat{\text{var}}(\hat{\theta}_1) = \frac{1}{N^{(T)}l_1^{(T)}(1-l_1^{(T)})} + \frac{1}{N^{(C)}l_1^{(C)}(1-l_1^{(C)})},$$

$$\hat{\text{var}}(\hat{\theta}_2) = \frac{1}{N^{(T)}l_2^{(T)}(1-l_2^{(T)})} + \frac{1}{N^{(C)}l_2^{(C)}(1-l_2^{(C)})}, \text{ and}$$

$$\hat{\text{cov}}(\hat{\theta}_1, \hat{\theta}_2) = \frac{1}{N^{(T)}l_1^{(T)}(1-l_2^{(T)})} + \frac{1}{N^{(C)}l_1^{(C)}(1-l_2^{(C)})}.$$

Priors for Bayesian analyses

The true probabilities in the controls are assigned flat Dirichlet priors. For mutually exclusive outcomes:

$$(\pi_{k1}^{(C)},\ldots,\pi_{kM}^{(C)}) \sim \mathcal{D}(1,\ldots,1) \quad \text{(Appendix. 19)}$$

For outcomes with an is-subset-of relationship, we assign priors to the probabilities for the mutually exclusive categories implied by the outcomes:

$$(\gamma_{k1}^{(C)},\ldots,\gamma_{kM}^{(C)}) \sim \mathcal{D}(1,\ldots,1) \quad \text{(Appendix. 20)}$$

The true between-study means (the elements of $\boldsymbol{\Theta}$) are assigned independent normal priors

$$\theta_m \sim \mathcal{N}(0,10^5). \quad \text{(Appendix. 21)}$$

To assign priors for the covariance matrix $\mathbf{T}$ we use the factorization $\mathbf{T} = \text{diag}(\boldsymbol{\tau})\,\mathbf{R}\,\text{diag}(\boldsymbol{\tau})$. We assign independent uniform priors to the elements of $\boldsymbol{\tau}$:

$$\tau_m \sim \mathcal{U}(10^{-4},2) \quad \text{(Appendix. 22)}$$

The priors for $\mathbf{R}$ must guarantee that the matrix is positive definite with elements between -1 and 1. Further, because the posterior distributions for the elements of $\mathbf{R}$ with uninformative priors were very wide, we performed analyses with uninformative and relatively informative priors.

For $M = 2$

We have a single correlation coefficient $\rho_{12} = \rho$. As noninformative prior we chose the uniform

$$\rho \sim \mathcal{U}(-0.9999,\ 0.9999). \quad \text{(Appendix. 23)}$$

For outcomes with an is-subset-of relationship, it is likely that ρ takes mostly positive values. We therefore repeated analyses by assigning the following prior to the Fisher-transformed correlation coefficient:

$$\tanh(\rho) \sim \mathcal{N}(0.5493,\ 0.6633) \quad \text{(Appendix. 24)}$$

where the mean corresponds to a correlation coefficient of 0.50, and the variance is such that 75% of the drawn values are positive

$$0.5493 = \tanh^{-1}(0.50), \text{ and} \quad \text{(Appendix. 25)}$$

$$0.6633 = (\tanh(0.50)/\Phi^{-1}(0.25))^2, \text{ with} \quad \text{(Appendix. 26)}$$

$\Phi^{-1}(\cdot)$ being the inverse cumulative distribution function for the standard normal. For mutually exclusive outcomes we do not anticipating a specific sign for the correlation coefficient, and thus we used only uninformative priors.

For $M = 3$

We follow Lu and Ades[1] in factorizing $\mathbf{R}$ using the Cholesky decomposition for square symmetric matrices $\mathbf{R} = \mathbf{U}'\mathbf{U}$, and in assigning specially constructed priors to the elements of the upper triangular matrix $\mathbf{U}$ (this is the spherical parameterization of Pinheiro and Bates[2]):

$$\mathbf{U} = \begin{bmatrix} 1 & \cos(\phi_1) & \cos(\phi_2) \\ 0 & \sin(\phi_1) & \sin(\phi_2)\cos(\phi_3) \\ 0 & 0 & \sin(\phi_2)\sin(\phi_3) \end{bmatrix} \qquad \text{(Appendix. 27)}$$

Setting uniform independent priors for ϕ_l $(l = 1, 2, 3)$ in the interval 0 to $\pi = 3.14159...$ is an option for obtaining a prior for $\mathbf{R}$ in which all elements are between -1 and 1 and positive definitiveness is guaranteed

$$\phi_l \sim \mathcal{U}(0, \pi) \qquad \text{(Appendix. 28)}$$

To restrict the elements of $\mathbf{R}$ to positive values use:

$$\phi_l \sim \mathcal{U}(0, \pi/2)\,. \qquad \text{(Appendix. 29)}$$

See Lu and Ades[1] for a short discussion of the implied density of the elements of $\mathbf{R}$ using the priors above. Nevertheless, the examined priors for the elements of $\mathbf{R}$ do not result in appreciable differences in the posterior distributions for the meta-analysis means. See Pinheiro and Bates[2] for a discussion of additional parameterizations.

Fitting the normal approximation model when at least one study has a singular covariance matrix

It is possible that the calculated covariance matrix of study k is singular. In a bivariate meta-analysis a study will have a singular covariance matrix if $p_{k,1}^{(T)} = p_{k,1}^{(C)} = p_{k,2}^{(T)} = p_{k,2}^{(C)} = 0.5$ for mutually exclusive outcomes; and if $l_{k,1}^{(T)} = l_{k,2}^{(T)}$ and $l_{k,1}^{(C)} = l_{k,2}^{(C)}$ for outcomes with an is-subset-of relationship. Singular covariance matrices cannot be inverted.[a] To make calculations possible, one option is to keep only one outcome in the offending study or studies. Another option is to keep all outcomes and to use a ridge regression approach. Ridge regression is a form of regularized regression where a positive number $\varepsilon_{\text{ridge}}$ (regularizing coefficient or regularizer) is added to the eigenvalues of the singular covariance matrices of offending studies.

Let $\mathbf{W}$ be the inverse of the covariance matrix of the multivariate meta-analysis summary log odds ratios Θ with fixed effects. Then Θ, and the determinant $\mathrm{D}(\varepsilon_{\text{ridge}})$ and the trace $\mathrm{Tr}(\varepsilon_{\text{ridge}})$ of $\mathbf{W}$ are functions of $\varepsilon_{\text{ridge}}$. One selects the minimum $\varepsilon_{\text{ridge}}$ so that $\mathrm{D}(\varepsilon_{\text{ridge}})$, $\mathrm{Tr}(\varepsilon_{\text{ridge}})$ and the elements of Θ are stable.

For example, in topic 2 in the main report (comparison between sirolimus-eluting and bare metal stents for acute cardiac disease) the STRATEGY trial has a singular covariance matrix. The outcomes of interest are the number of major cardiac events at 6 and 12 months. No additional events were observed between 6 and 12 months in either arm, and $l_{\text{STRATEGY},1}^{(T)} = l_{\text{STRATEGY},2}^{(T)}$ and $l_{\text{STRATEGY},1}^{(C)} = l_{\text{STRATEGY},2}^{(C)}$. The (uncorrected) sampling covariance matrix is calculated as

[a] A similar concern pertains to studies with near singular covariance matrices, in that their inverses are numerically unstable. The same regularized regression can be used to correct near singular matrices. In this work we did not check for near-singular matrices nor made any corrections for them.

$$\mathbf{S}_{\text{STRATEGY}} = \begin{bmatrix} .1289655 & .1289655 \\ & .1289655 \end{bmatrix},$$

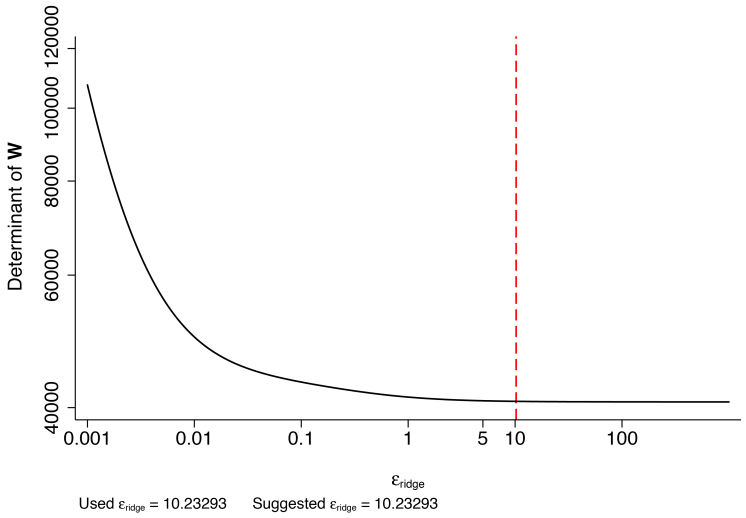
and is singular. The corrected covariance matrix $\mathbf{S}_{\text{STRATEGY}}(\varepsilon_{\text{ridge}})$ is obtained by adding $\varepsilon_{\text{ridge}}$ to the eigenvalues of $\mathbf{S}_{\text{STRATEGY}}$. It is easy to show that

$$\mathbf{S}_{\text{STRATEGY}}(\varepsilon_{\text{ridge}}) = \mathbf{S}_{\text{STRATEGY}} + \mathbf{I}\varepsilon_{\text{ridge}} = \begin{bmatrix} .1289655+\varepsilon_{\text{ridge}} & .1289655 \\ & .1289655+\varepsilon_{\text{ridge}} \end{bmatrix},$$

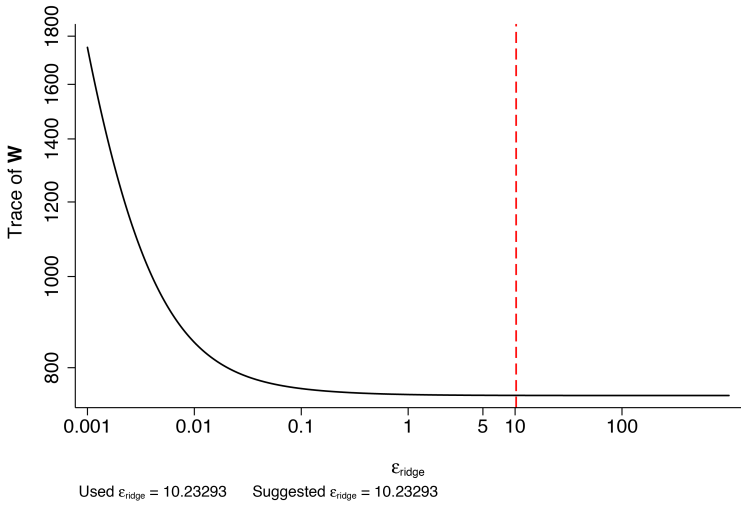
where $\mathbf{I}$ is the identity matrix. We examine the plots of $\mathrm{D}(\varepsilon_{\text{ridge}})$ (Appendix Figure 1), $\mathrm{Tr}(\varepsilon_{\text{ridge}})$ (Appendix Figure 2), and Θ (Appendix Figure 3) to select a value for $\varepsilon_{\text{ridge}}$. An infinitely large $\varepsilon_{\text{ridge}}$ is equivalent to excluding STRATEGY from both outcomes in the meta-analysis. The aim is to select the minimum value for which the quantities in the figures have stabilized.

We devised a heuristic for selecting the value of $\varepsilon_{\text{ridge}}$ to enhance consistency in calculations across topics. We calculated the average of the absolute values of the smoothed numerical derivatives (slopes) of $\mathrm{D}(\varepsilon_{\text{ridge}})$, $\mathrm{Tr}(\varepsilon_{\text{ridge}})$, and of each summary log odds ratio with respect to ε. The heuristic proposes to use the ε that corresponds to an average absolute slope 1.7, i.e., a tangent angle of 60 degrees (Appendix Figure 4).

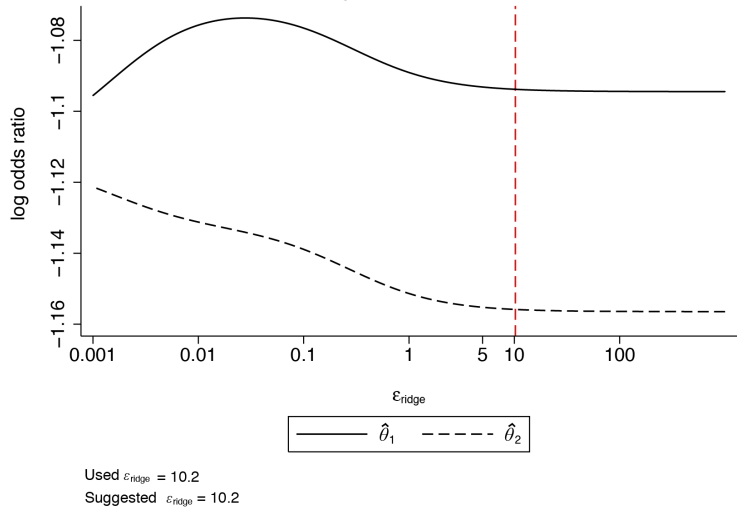
Appendix Figure 1. Relationship between the determinant of the inverse covariance matrix of the fixed effects multivariate meta-analysis estimates $\mathbf{W}$ as a function of $\varepsilon_{\text{ridge}}$ (topic 2)



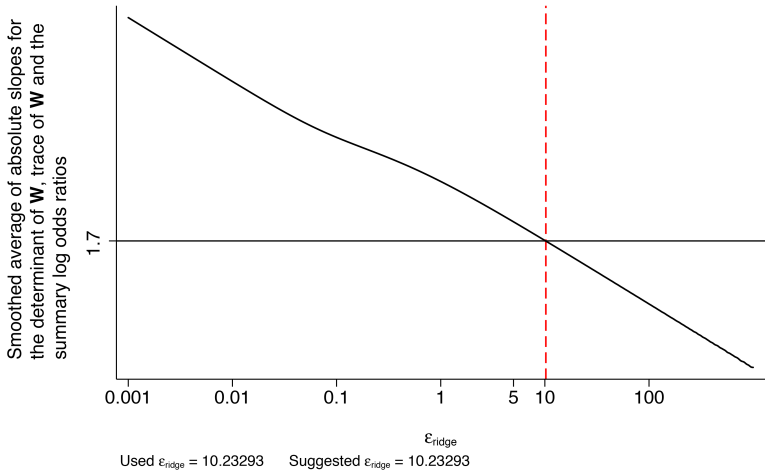
Appendix Figure 2. Relationship between the trace of the inverse covariance matrix of the fixed effects multivariate meta-analysis estimates $\mathbf{W}$ as a function of $\varepsilon_{\text{ridge}}$ (topic 2)



Appendix Figure 3. Relationship between the summary log odds ratios from fixed effects multivariate meta-analysis as a function of $\varepsilon_{\text{ridge}}$ (topic 2)



Appendix Figure 4. Heuristic for picking the value of $\varepsilon_{\text{ridge}}$ (topic 2)



Additional Results (Tables)

Appendix Table 1. Estimates with univariate and multivariate meta-analysis using normal approximation models in topics 1 through 43 (random effects)

Topic	Univariate $\exp(\hat{\theta}_1)$	Univariate $\exp(\hat{\theta}_2)$	Univariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1)$	Multivariate $\exp(\hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$
1	0.591 (0.484, 0.723)	0.651 (0.472, 0.899)	0.908 (0.620, 1.331)	0.590 (0.496, 0.702)	0.666 (0.517, 0.859)	0.885 (0.730, 1.074)
2	0.294 (0.229, 0.377)	0.339 (0.264, 0.435)	0.868 (0.610, 1.235)	0.324 (0.249, 0.421)	0.303 (0.235, 0.389)	1.071 (0.954, 1.201)
3	1.833 (1.354, 2.481)	1.618 (1.099, 2.383)	1.133 (0.692, 1.855)	1.831 (1.208, 2.774)	1.460 (1.007, 2.116)	1.254 (0.985, 1.596)
4	1.132 (0.692, 1.852)	0.938 (0.664, 1.326)	1.206 (0.661, 2.202)	1.162 (0.690, 1.957)	0.942 (0.650, 1.365)	1.233 (0.757, 2.010)
5	0.788 (0.483, 1.285)	0.789 (0.531, 1.174)	0.998 (0.532, 1.873)	0.851 (0.545, 1.330)	0.761 (0.495, 1.171)	1.118 (0.618, 2.020)
6	0.931 (0.569, 1.522)	0.885 (0.612, 1.281)	1.051 (0.568, 1.947)	0.966 (0.606, 1.542)	0.876 (0.601, 1.277)	1.103 (0.802, 1.517)
7	0.486 (0.299, 0.791)	0.354 (0.215, 0.583)	1.373 (0.684, 2.755)	0.510 (0.313, 0.832)	0.373 (0.230, 0.606)	1.368 (0.880, 2.127)
8	0.666 (0.298, 1.492)	0.845 (0.655, 1.091)	0.788 (0.338, 1.840)	0.699 (0.303, 1.613)	0.855 (0.651, 1.123)	0.818 (0.373, 1.795)
9	1.922 (1.402, 2.634)	1.021 (0.848, 1.231)	1.881 (1.303, 2.716)	1.794 (1.301, 2.475)	1.055 (0.867, 1.283)	1.701 (1.252, 2.312)
10	0.871 (0.762, 0.995)	0.836 (0.745, 0.937)	1.042 (0.873, 1.243)	0.822 (0.732, 0.923)	0.839 (0.751, 0.937)	0.980 (0.925, 1.038)
11	1.003 (0.762, 1.320)	1.137 (0.665, 1.945)	0.882 (0.482, 1.612)	0.990 (0.733, 1.337)	1.092 (0.667, 1.785)	0.907 (0.692, 1.190)
12	0.585 (0.430, 0.796)	0.629 (0.451, 0.878)	0.930 (0.589, 1.469)	0.622 (0.463, 0.834)	0.627 (0.468, 0.842)	0.991 (0.883, 1.111)
13	0.790 (0.640, 0.974)	0.801 (0.693, 0.926)	0.985 (0.763, 1.273)	0.795 (0.657, 0.961)	0.830 (0.708, 0.973)	0.958 (0.837, 1.095)
14	0.809 (0.607, 1.077)	0.852 (0.657, 1.106)	0.949 (0.644, 1.397)	0.853 (0.639, 1.138)	0.912 (0.711, 1.169)	0.935 (0.769, 1.138)
15	0.377 (0.162, 0.880)	0.706 (0.461, 1.079)	0.535 (0.206, 1.386)	0.389 (0.163, 0.927)	0.673 (0.456, 0.994)	0.577 (0.311, 1.072)
16	0.386 (0.262, 0.568)	0.530 (0.443, 0.633)	0.728 (0.474, 1.117)	0.421 (0.317, 0.560)	0.529 (0.444, 0.631)	0.796 (0.677, 0.936)
17	0.342 (0.198, 0.589)	0.517 (0.367, 0.728)	0.661 (0.346, 1.263)	0.326 (0.185, 0.576)	0.519 (0.374, 0.720)	0.629 (0.392, 1.011)
18	0.358 (0.240, 0.534)	0.281 (0.216, 0.366)	1.273 (0.785, 2.062)	0.436 (0.311, 0.610)	0.281 (0.216, 0.366)	1.552 (1.238, 1.946)
19	11.163 (5.810, 21.447)	1.943 (0.683, 5.534)	5.744 (1.651, 19.989)	10.932 (5.797, 20.615)	1.848 (0.666, 5.123)	5.917 (2.034, 17.212)
20	5.529 (2.099, 14.569)	3.296 (1.738, 6.254)	1.677 (0.522, 5.394)	5.668 (2.515, 12.774)	3.670 (1.782, 7.561)	1.544 (0.761, 3.132)
21	0.722 (0.572, 0.912)	0.568 (0.486, 0.663)	1.273 (0.961, 1.685)	0.780 (0.613, 0.992)	0.551 (0.464, 0.654)	1.415 (1.079, 1.856)
22	0.985 (0.608, 1.594)	1.096 (0.721, 1.667)	0.899 (0.474, 1.702)	0.964 (0.575, 1.615)	1.161 (0.701, 1.924)	0.830 (0.636, 1.083)
23	1.065 (0.632, 1.794)	0.875 (0.643, 1.190)	1.218 (0.665, 2.231)	1.032 (0.573, 1.860)	0.868 (0.629, 1.198)	1.190 (0.672, 2.107)
24	0.753 (0.366, 1.552)	0.815 (0.428, 1.551)	0.925 (0.351, 2.434)	0.786 (0.340, 1.818)	0.740 (0.343, 1.598)	1.062 (0.729, 1.546)
25	1.269 (0.576, 2.799)	1.848 (1.060, 3.222)	0.687 (0.261, 1.811)	1.201 (0.601, 2.402)	1.979 (1.061, 3.693)	0.607 (0.358, 1.030)
26	0.467 (0.404, 0.540)	0.315 (0.266, 0.374)	1.481 (1.183, 1.854)	0.502 (0.443, 0.570)	0.347 (0.299, 0.401)	1.450 (1.231, 1.707)
27	0.452 (0.398, 0.513)	0.397 (0.344, 0.459)	1.137 (0.938, 1.378)	0.427 (0.367, 0.497)	0.379 (0.317, 0.454)	1.126 (0.918, 1.380)
28	0.342 (0.290, 0.403)	0.278 (0.229, 0.338)	1.228 (0.952, 1.585)	0.352 (0.299, 0.414)	0.272 (0.228, 0.325)	1.292 (1.113, 1.500)
29	0.617 (0.519, 0.734)	0.544 (0.461, 0.641)	1.136 (0.895, 1.442)	0.638 (0.542, 0.751)	0.560 (0.478, 0.656)	1.140 (1.024, 1.268)
30	0.237 (0.194, 0.289)	0.244 (0.200, 0.298)	0.971 (0.732, 1.288)	0.248 (0.204, 0.302)	0.176 (0.138, 0.224)	1.412 (1.190, 1.676)
31	0.443 (0.333, 0.589)	0.442 (0.320, 0.610)	1.002 (0.651, 1.544)	0.473 (0.364, 0.615)	0.461 (0.344, 0.618)	1.026 (0.781, 1.348)
32	0.637 (0.517, 0.784)	0.410 (0.289, 0.580)	1.554 (1.035, 2.334)	0.653 (0.520, 0.821)	0.423 (0.271, 0.661)	1.544 (0.951, 2.507)

Appendix Table1. Estimates with univariate and multivariate meta-analysis using normal approximation models in topics 1 through 43 (random effects) (continued)

Topic	Univariate $\exp(\hat{\theta}_1)$	Univariate $\exp(\hat{\theta}_2)$	Univariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1)$	Multivariate $\exp(\hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$
33	0.418 (0.241, 0.723)	0.458 (0.264, 0.794)	0.913 (0.416, 2.003)	0.410 (0.241, 0.698)	0.314 (0.186, 0.532)	1.305 (0.788, 2.160)
34	2.921 (2.102, 4.060)	3.060 (1.991, 4.702)	0.955 (0.554, 1.644)	2.953 (2.081, 4.192)	3.544 (2.307, 5.443)	0.833 (0.574, 1.210)
35	0.810 (0.574, 1.143)	0.589 (0.428, 0.811)	1.374 (0.858, 2.201)	0.813 (0.571, 1.157)	0.568 (0.413, 0.782)	1.431 (1.003, 2.041)
36	0.289 (0.161, 0.516)	0.208 (0.118, 0.367)	1.387 (0.613, 3.137)	0.298 (0.170, 0.522)	0.192 (0.114, 0.325)	1.550 (1.012, 2.375)
37	1.233 (0.605, 2.513)	1.414 (0.780, 2.565)	0.872 (0.345, 2.206)	1.079 (0.568, 2.050)	1.433 (0.817, 2.512)	0.753 (0.468, 1.212)
38	0.676 (0.431, 1.060)	0.661 (0.493, 0.886)	1.023 (0.596, 1.755)	0.583 (0.399, 0.852)	0.624 (0.446, 0.872)	0.934 (0.771, 1.132)
39*	1.923 (0.986, 3.751)	0.542 (0.351, 0.838)	3.546 (1.594, 7.888)	1.726 (0.995, 2.994)	0.531 (0.337, 0.836)	3.250 (1.258, 8.401)
40*	0.792 (0.453, 1.385)	0.709 (0.417, 1.205)	1.117 (0.517, 2.414)	0.830 (0.477, 1.442)	0.671 (0.389, 1.157)	1.236 (0.626, 2.441)
41*	1.223 (1.069, 1.398)	0.891 (0.764, 1.039)	1.373 (1.118, 1.685)	1.261 (1.071, 1.485)	0.843 (0.711, 0.998)	1.497 (1.099, 2.038)
42*	0.908 (0.706, 1.167)	0.972 (0.759, 1.244)	0.934 (0.655, 1.332)	0.869 (0.631, 1.197)	0.950 (0.720, 1.252)	0.915 (0.619, 1.352)
43*	1.211 (1.077, 1.362)	0.921 (0.780, 1.087)	1.315 (1.069, 1.618)	1.217 (1.086, 1.365)	0.929 (0.798, 1.082)	1.310 (1.033, 1.661)

Shown are means and 95% confidence intervals.
*These topics refer to mutually exclusive outcomes.

Appendix Table 2. Estimates with univariate and multivariate meta-analysis using normal approximation models in topics 44 and 45 (random effects)

Analysis or Topic	$\exp\left(\hat{\theta}_1\right)$	$\exp\left(\hat{\theta}_2\right)$	$\exp\left(\hat{\theta}_3\right)$	$\exp\left(\hat{\theta}_1 - \hat{\theta}_2\right)$	$\exp\left(\hat{\theta}_1 - \hat{\theta}_3\right)$	$\exp\left(\hat{\theta}_2 - \hat{\theta}_3\right)$
Univariate						
44	1.957 (1.201, 3.189)	1.620 (1.089, 2.411)	1.586 (1.073, 2.343)	1.208 (0.644, 2.267)	1.234 (0.660, 2.306)	1.022 (0.585, 1.784)
45	1.195 (0.601, 2.374)	1.137 (0.704, 1.837)	1.058 (0.719, 1.557)	1.050 (0.451, 2.445)	1.129 (0.512, 2.488)	1.075 (0.577, 2.002)
Multivariate						
44	1.653 (1.107, 2.467)	1.941 (1.202, 3.133)	1.672 (1.091, 2.563)	0.852 (0.564, 1.286)	0.988 (0.674, 1.450)	1.161 (0.758, 1.777)
45	1.034 (0.673, 1.590)	1.099 (0.613, 1.968)	1.067 (0.696, 1.637)	0.941 (0.739, 1.199)	0.969 (0.872, 1.076)	1.029 (0.780, 1.358)

Shown are means and 95% confidence intervals.

Appendix Table 3. Estimates with univariate and multivariate meta-analysis using normal approximation models in topics 1 through 43 (fixed effects)

Topic	Univariate $\exp(\hat{\theta}_1)$	Univariate $\exp(\hat{\theta}_2)$	Univariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1)$	Multivariate $\exp(\hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$
1	0.591 (0.498, 0.702)	0.643 (0.529, 0.781)	0.920 (0.708, 1.195)	0.590 (0.501, 0.696)	0.650 (0.553, 0.765)	0.908 (0.823, 1.002)
2	0.294 (0.235, 0.368)	0.344 (0.279, 0.423)	0.856 (0.629, 1.163)	0.343 (0.282, 0.419)	0.321 (0.266, 0.388)	1.070 (0.956, 1.198)
3	1.544 (1.393, 1.712)	1.297 (1.186, 1.419)	1.190 (1.038, 1.365)	1.545 (1.395, 1.710)	1.291 (1.181, 1.411)	1.197 (1.106, 1.295)
4	1.132 (0.692, 1.852)	0.938 (0.664, 1.326)	1.206 (0.661, 2.202)	1.114 (0.688, 1.806)	0.943 (0.667, 1.334)	1.181 (0.778, 1.793)
5	0.763 (0.548, 1.062)	0.789 (0.552, 1.129)	0.966 (0.594, 1.573)	0.783 (0.563, 1.088)	0.804 (0.566, 1.143)	0.973 (0.669, 1.416)
6	0.931 (0.569, 1.522)	0.886 (0.614, 1.277)	1.051 (0.569, 1.943)	0.968 (0.615, 1.525)	0.877 (0.607, 1.265)	1.105 (0.805, 1.517)
7	0.486 (0.299, 0.791)	0.352 (0.222, 0.558)	1.380 (0.707, 2.696)	0.509 (0.315, 0.822)	0.371 (0.235, 0.583)	1.373 (0.892, 2.114)
8	0.716 (0.362, 1.418)	0.845 (0.655, 1.091)	0.847 (0.408, 1.760)	0.751 (0.393, 1.435)	0.846 (0.656, 1.092)	0.887 (0.478, 1.648)
9	1.922 (1.402, 2.634)	1.018 (0.851, 1.218)	1.888 (1.313, 2.714)	1.679 (1.247, 2.259)	1.032 (0.863, 1.234)	1.627 (1.256, 2.107)
10	0.871 (0.762, 0.995)	0.836 (0.745, 0.937)	1.042 (0.873, 1.243)	0.817 (0.729, 0.915)	0.841 (0.754, 0.938)	0.971 (0.928, 1.017)
11	0.984 (0.826, 1.172)	1.184 (0.955, 1.468)	0.831 (0.629, 1.097)	0.986 (0.828, 1.173)	1.123 (0.915, 1.378)	0.878 (0.741, 1.041)
12	0.655 (0.532, 0.807)	0.729 (0.600, 0.885)	0.899 (0.674, 1.199)	0.695 (0.569, 0.849)	0.709 (0.589, 0.854)	0.980 (0.875, 1.098)
13	0.788 (0.691, 0.900)	0.801 (0.693, 0.926)	0.984 (0.808, 1.198)	0.784 (0.691, 0.891)	0.844 (0.739, 0.963)	0.929 (0.844, 1.023)
14	0.778 (0.649, 0.932)	0.857 (0.721, 1.018)	0.908 (0.707, 1.166)	0.820 (0.696, 0.967)	0.914 (0.779, 1.073)	0.897 (0.798, 1.010)
15	0.935 (0.712, 1.228)	0.932 (0.790, 1.099)	1.003 (0.728, 1.382)	0.880 (0.677, 1.144)	0.934 (0.792, 1.102)	0.942 (0.751, 1.182)
16	0.398 (0.333, 0.476)	0.524 (0.471, 0.583)	0.760 (0.616, 0.937)	0.417 (0.355, 0.490)	0.525 (0.472, 0.584)	0.794 (0.699, 0.902)
17	0.437 (0.331, 0.576)	0.590 (0.495, 0.705)	0.740 (0.532, 1.029)	0.452 (0.352, 0.579)	0.598 (0.502, 0.714)	0.755 (0.626, 0.909)
18	0.358 (0.240, 0.534)	0.281 (0.216, 0.366)	1.273 (0.785, 2.062)	0.436 (0.311, 0.610)	0.281 (0.216, 0.366)	1.552 (1.238, 1.946)
19	11.163 (5.810, 21.447)	1.943 (0.683, 5.534)	5.744 (1.651, 19.989)	10.932 (5.797, 20.615)	1.848 (0.666, 5.123)	5.917 (2.034, 17.212)
20	4.051 (2.483, 6.609)	3.272 (1.853, 5.780)	1.238 (0.580, 2.643)	4.348 (2.715, 6.964)	3.271 (1.859, 5.755)	1.329 (0.750, 2.355)
21	0.722 (0.572, 0.912)	0.585 (0.516, 0.663)	1.235 (0.948, 1.609)	0.793 (0.633, 0.993)	0.585 (0.516, 0.663)	1.356 (1.101, 1.670)
22	0.985 (0.608, 1.594)	1.096 (0.721, 1.667)	0.899 (0.474, 1.702)	0.928 (0.580, 1.483)	1.109 (0.729, 1.688)	0.836 (0.657, 1.065)
23	1.059 (0.662, 1.695)	0.880 (0.663, 1.170)	1.203 (0.695, 2.084)	0.848 (0.546, 1.316)	0.888 (0.668, 1.180)	0.955 (0.671, 1.358)
24	0.753 (0.366, 1.552)	0.815 (0.428, 1.551)	0.925 (0.351, 2.434)	0.862 (0.434, 1.710)	0.792 (0.414, 1.516)	1.088 (0.783, 1.512)
25	1.239 (0.655, 2.344)	1.644 (1.123, 2.406)	0.753 (0.358, 1.586)	1.081 (0.628, 1.862)	1.708 (1.174, 2.485)	0.633 (0.401, 0.999)
26	0.495 (0.460, 0.533)	0.298 (0.266, 0.333)	1.665 (1.457, 1.902)	0.472 (0.439, 0.507)	0.402 (0.367, 0.441)	1.173 (1.092, 1.261)
27	0.463 (0.413, 0.519)	0.433 (0.384, 0.487)	1.070 (0.908, 1.261)	0.432 (0.390, 0.479)	0.437 (0.393, 0.485)	0.990 (0.940, 1.043)
28	0.346 (0.298, 0.402)	0.283 (0.240, 0.334)	1.223 (0.979, 1.529)	0.359 (0.311, 0.415)	0.279 (0.241, 0.322)	1.289 (1.122, 1.481)
29	0.617 (0.519, 0.734)	0.544 (0.461, 0.641)	1.136 (0.895, 1.442)	0.613 (0.527, 0.714)	0.544 (0.467, 0.633)	1.128 (1.015, 1.254)
30	0.248 (0.209, 0.294)	0.244 (0.200, 0.298)	1.017 (0.782, 1.324)	0.265 (0.225, 0.312)	0.197 (0.166, 0.235)	1.343 (1.156, 1.560)
31	0.471 (0.377, 0.589)	0.461 (0.356, 0.596)	1.023 (0.728, 1.439)	0.496 (0.403, 0.611)	0.484 (0.384, 0.611)	1.025 (0.803, 1.307)
32	0.637 (0.517, 0.784)	0.444 (0.339, 0.582)	1.434 (1.019, 2.018)	0.595 (0.486, 0.729)	0.512 (0.400, 0.655)	1.162 (0.938, 1.438)
33	0.531 (0.397, 0.710)	0.569 (0.398, 0.814)	0.933 (0.586, 1.484)	0.539 (0.404, 0.719)	0.373 (0.245, 0.568)	1.445 (0.944, 2.212)
34	2.921 (2.102, 4.060)	2.604 (1.843, 3.678)	1.122 (0.695, 1.812)	2.988 (2.156, 4.141)	2.839 (2.066, 3.902)	1.052 (0.808, 1.372)
35	0.810 (0.574, 1.143)	0.589 (0.428, 0.811)	1.374 (0.858, 2.201)	0.789 (0.567, 1.097)	0.560 (0.408, 0.770)	1.408 (0.994, 1.993)

Appendix Table 3. Estimates with univariate and multivariate meta-analysis using normal approximation models in topics 1 through 43 (fixed effects) (continued)

Topic	Univariate $\exp(\hat{\theta}_1)$	Univariate $\exp(\hat{\theta}_2)$	Univariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1)$	Multivariate $\exp(\hat{\theta}_2)$	Multivariate $\exp(\hat{\theta}_1 - \hat{\theta}_2)$
36	0.289 (0.161, 0.516)	0.208 (0.118, 0.367)	1.387 (0.613, 3.137)	0.298 (0.170, 0.522)	0.192 (0.114, 0.325)	1.550 (1.012, 2.375)
37	1.233 (0.605, 2.513)	1.414 (0.800, 2.500)	0.872 (0.350, 2.170)	1.079 (0.568, 2.050)	1.433 (0.817, 2.512)	0.753 (0.468, 1.212)
38	0.676 (0.435, 1.051)	0.661 (0.493, 0.886)	1.023 (0.601, 1.742)	0.649 (0.471, 0.894)	0.678 (0.503, 0.913)	0.957 (0.794, 1.153)
39*	1.561 (1.213, 2.009)	0.500 (0.375, 0.667)	3.121 (2.124, 4.585)	1.539 (1.203, 1.969)	0.529 (0.398, 0.702)	2.911 (1.876, 4.518)
40*	0.887 (0.647, 1.216)	0.702 (0.445, 1.109)	1.263 (0.725, 2.200)	0.945 (0.691, 1.292)	0.829 (0.529, 1.299)	1.139 (0.599, 2.168)
41*	1.141 (1.050, 1.241)	0.924 (0.827, 1.032)	1.235 (1.075, 1.421)	1.150 (1.059, 1.248)	0.903 (0.814, 1.001)	1.274 (1.075, 1.510)
42*	0.902 (0.732, 1.111)	0.959 (0.767, 1.198)	0.941 (0.692, 1.279)	0.904 (0.734, 1.114)	0.961 (0.769, 1.201)	0.941 (0.690, 1.281)
43*	1.211 (1.077, 1.362)	0.921 (0.790, 1.074)	1.315 (1.081, 1.601)	1.217 (1.086, 1.365)	0.929 (0.798, 1.082)	1.310 (1.033, 1.661)

Shown are means and 95% confidence intervals.
*These topics refer to mutually exclusive outcomes.

Appendix Table 4. Estimates with univariate and multivariate meta-analysis using normal approximation models in topics 44 and 45 (fixed effects)

Analysis or Topic	$\exp\left(\hat{\theta}_1\right)$	$\exp\left(\hat{\theta}_2\right)$	$\exp\left(\hat{\theta}_3\right)$	$\exp\left(\hat{\theta}_1 - \hat{\theta}_2\right)$	$\exp\left(\hat{\theta}_1 - \hat{\theta}_3\right)$	$\exp\left(\hat{\theta}_2 - \hat{\theta}_3\right)$
Univariate						
44	1.957 (1.201, 3.189)	1.620 (1.089, 2.411)	1.586 (1.073, 2.343)	1.208 (0.644, 2.267)	1.234 (0.660, 2.306)	1.022 (0.585, 1.784)
45	1.355 (1.112, 1.651)	1.201 (1.034, 1.394)	1.200 (1.044, 1.378)	1.128 (0.879, 1.448)	1.129 (0.886, 1.439)	1.001 (0.814, 1.230)
Multivariate						
44	1.706 (1.164, 2.499)	1.766 (1.142, 2.733)	1.683 (1.153, 2.457)	0.966 (0.721, 1.293)	1.014 (0.802, 1.281)	1.049 (0.728, 1.513)
45	1.190 (1.031, 1.375)	1.327 (1.090, 1.616)	1.218 (1.061, 1.399)	0.897 (0.773, 1.040)	0.977 (0.941, 1.014)	1.089 (0.935, 1.268)

Shown are means and 95% confidence intervals.

Appendix Table 5. Estimates with univariate and multivariate meta-analysis using the binomial or multinomial likelihood in topics 1 through 43 (random effects)

Topic	Univariate $\exp(\theta_1)$	Univariate $\exp(\theta_2)$	Univariate $\exp(\theta_1 - \theta_2)$	Multivariate $\exp(\theta_1)$	Multivariate $\exp(\theta_2)$	Multivariate $\exp(\theta_1 - \theta_2)$
1	0.545 (0.382, 0.684)	0.586 (0.346, 0.909)	0.925 (0.526, 1.658)	0.540 (0.409, 0.650)	0.612 (0.448, 0.806)	0.877 (0.693, 1.111)
2	0.259 (0.178, 0.344)	0.261 (0.170, 0.368)	0.993 (0.600, 1.648)	0.287 (0.215, 0.368)	0.301 (0.234, 0.373)	0.964 (0.753, 1.131)
3	1.748 (1.356, 2.719)	1.533 (1.145, 2.600)	1.141 (0.633, 1.960)	1.659 (1.312, 2.455)	1.386 (1.066, 2.018)	1.197 (0.877, 1.688)
4	0.934 (0.463, 1.605)	0.767 (0.470, 1.143)	1.220 (0.553, 2.534)	0.851 (0.458, 1.451)	0.793 (0.523, 1.168)	1.072 (0.655, 1.681)
5	0.771 (0.455, 1.284)	0.712 (0.400, 1.096)	1.093 (0.547, 2.375)	0.785 (0.467, 1.308)	0.759 (0.432, 1.148)	1.035 (0.610, 1.876)
6	0.602 (0.245, 1.067)	0.495 (0.191, 0.816)	1.234 (0.441, 3.663)	0.520 (0.215, 0.915)	0.572 (0.306, 0.923)	0.889 (0.537, 1.281)
7	0.480 (0.275, 0.799)	0.351 (0.205, 0.598)	1.365 (0.644, 2.863)	0.471 (0.278, 0.809)	0.377 (0.223, 0.653)	1.249 (0.713, 2.115)
8	0.477 (0.155, 1.100)	0.818 (0.613, 1.112)	0.577 (0.186, 1.413)	0.474 (0.147, 1.139)	0.814 (0.615, 1.126)	0.579 (0.190, 1.340)
9	1.512 (1.090, 2.049)	0.857 (0.697, 1.038)	1.764 (1.215, 2.533)	1.320 (0.977, 1.812)	0.862 (0.716, 1.048)	1.524 (1.186, 2.016)
10	0.869 (0.753, 1.004)	0.839 (0.725, 0.976)	1.035 (0.844, 1.273)	0.795 (0.713, 0.878)	0.802 (0.716, 0.872)	0.994 (0.938, 1.053)
11	1.044 (0.781, 1.410)	1.147 (0.643, 1.959)	0.912 (0.495, 1.717)	1.016 (0.757, 1.366)	1.089 (0.646, 1.772)	0.933 (0.682, 1.323)
12	0.528 (0.365, 0.722)	0.549 (0.351, 0.774)	0.964 (0.581, 1.649)	0.539 (0.376, 0.700)	0.591 (0.423, 0.751)	0.920 (0.766, 1.060)
13	0.797 (0.635, 0.995)	0.793 (0.657, 0.970)	1.002 (0.742, 1.343)	0.792 (0.651, 0.964)	0.822 (0.705, 0.963)	0.961 (0.816, 1.148)
14	0.822 (0.601, 1.169)	0.776 (0.590, 1.088)	1.060 (0.668, 1.651)	0.854 (0.650, 1.156)	0.908 (0.707, 1.188)	0.938 (0.746, 1.208)
15	0.295 (0.094, 0.727)	0.715 (0.401, 1.161)	0.412 (0.121, 1.182)	0.280 (0.102, 0.655)	0.640 (0.404, 0.984)	0.444 (0.185, 0.874)
16	0.361 (0.227, 0.559)	0.489 (0.384, 0.627)	0.736 (0.437, 1.207)	0.371 (0.260, 0.510)	0.500 (0.415, 0.594)	0.744 (0.574, 0.928)
17	0.276 (0.116, 0.522)	0.513 (0.329, 0.752)	0.539 (0.210, 1.144)	0.222 (0.089, 0.423)	0.428 (0.274, 0.616)	0.519 (0.248, 0.901)
18	0.305 (0.180, 0.470)	0.223 (0.143, 0.320)	1.378 (0.722, 2.518)	0.324 (0.219, 0.465)	0.244 (0.183, 0.323)	1.327 (0.955, 1.769)
19	17.187 (7.959, 45.174)	40.932 (4.547, 1710.109)	0.424 (0.009, 4.781)	14.565 (6.902, 34.201)	13.113 (2.646, 123.898)	1.120 (0.124, 5.449)
20	7.209 (2.704, 21.568)	4.874 (1.931, 16.060)	1.477 (0.327, 6.239)	7.305 (3.179, 19.494)	5.967 (2.770, 17.124)	1.205 (0.513, 2.701)
21	0.596 (0.438, 0.761)	0.514 (0.415, 0.602)	1.163 (0.816, 1.598)	0.599 (0.432, 0.767)	0.520 (0.426, 0.610)	1.158 (0.829, 1.519)
22	0.789 (0.464, 1.262)	0.797 (0.489, 1.234)	0.991 (0.503, 1.926)	0.708 (0.449, 1.098)	0.818 (0.542, 1.197)	0.875 (0.656, 1.118)
23	0.845 (0.484, 1.406)	0.699 (0.478, 0.956)	1.209 (0.635, 2.293)	0.739 (0.404, 1.248)	0.699 (0.488, 0.948)	1.050 (0.653, 1.680)
24	0.467 (0.158, 1.010)	0.433 (0.179, 0.839)	1.079 (0.309, 3.452)	0.405 (0.167, 0.792)	0.462 (0.228, 0.817)	0.891 (0.480, 1.302)
25	1.096 (0.444, 2.499)	1.460 (0.736, 3.251)	0.743 (0.229, 2.182)	1.060 (0.537, 2.117)	1.634 (1.026, 2.979)	0.635 (0.348, 1.125)
26	0.474 (0.407, 0.550)	0.301 (0.246, 0.367)	1.572 (1.226, 2.019)	0.403 (0.351, 0.462)	0.380 (0.330, 0.436)	1.054 (1.001, 1.157)
27	0.448 (0.387, 0.511)	0.380 (0.307, 0.459)	1.177 (0.928, 1.517)	0.401 (0.351, 0.449)	0.394 (0.342, 0.444)	1.013 (0.961, 1.105)
28	0.314 (0.264, 0.367)	0.243 (0.198, 0.294)	1.293 (0.998, 1.680)	0.302 (0.255, 0.354)	0.263 (0.223, 0.314)	1.148 (0.978, 1.332)
29	0.626 (0.527, 0.745)	0.515 (0.424, 0.623)	1.216 (0.940, 1.572)	0.564 (0.488, 0.657)	0.542 (0.469, 0.627)	1.039 (0.977, 1.132)
30	0.225 (0.181, 0.271)	0.203 (0.157, 0.251)	1.111 (0.819, 1.519)	0.215 (0.177, 0.257)	0.188 (0.153, 0.228)	1.149 (0.957, 1.356)
31	0.434 (0.308, 0.586)	0.336 (0.227, 0.509)	1.287 (0.762, 2.096)	0.449 (0.331, 0.578)	0.441 (0.324, 0.590)	1.018 (0.764, 1.326)
32	0.611 (0.443, 0.782)	0.373 (0.220, 0.545)	1.640 (0.989, 2.873)	0.559 (0.381, 0.710)	0.459 (0.298, 0.620)	1.197 (0.936, 1.721)
33	0.415 (0.220, 0.719)	0.400 (0.179, 0.740)	1.045 (0.424, 2.763)	0.421 (0.232, 0.707)	0.389 (0.189, 0.671)	1.096 (0.611, 1.937)
34	2.535 (1.864, 3.497)	2.143 (1.530, 3.145)	1.182 (0.722, 1.903)	2.401 (1.776, 3.345)	2.610 (1.838, 3.906)	0.921 (0.659, 1.269)

Appendix Table 5. Estimates with univariate and multivariate meta-analysis using the binomial or multinomial likelihood in topics 1 through 43 (random effects) (continued)

Topic	Univariate $\exp(\theta_1)$	Univariate $\exp(\theta_2)$	Univariate $\exp(\theta_1-\theta_2)$	Multivariate $\exp(\theta_1)$	Multivariate $\exp(\theta_2)$	Multivariate $\exp(\theta_1-\theta_2)$
35	0.792 (0.562, 1.108)	0.577 (0.395, 0.840)	1.378 (0.825, 2.271)	0.637 (0.482, 0.853)	0.574 (0.425, 0.776)	1.112 (0.894, 1.441)
36	0.258 (0.142, 0.453)	0.160 (0.086, 0.280)	1.611 (0.716, 3.708)	0.218 (0.122, 0.375)	0.172 (0.103, 0.303)	1.229 (0.790, 1.942)
37	1.169 (0.546, 2.481)	1.263 (0.636, 2.507)	0.922 (0.330, 2.559)	1.019 (0.548, 2.061)	1.334 (0.745, 2.442)	0.772 (0.440, 1.377)
38	0.599 (0.299, 1.029)	0.457 (0.230, 0.774)	1.314 (0.552, 3.099)	0.501 (0.344, 0.684)	0.498 (0.359, 0.668)	1.001 (0.841, 1.193)
39*	1.895 (0.987, 3.858)	0.509 (0.336, 0.804)	3.717 (1.686, 8.401)	1.768 (0.980, 3.320)	0.503 (0.330, 0.783)	3.517 (1.417, 8.867)
40*	0.836 (0.453, 1.485)	0.624 (0.365, 1.075)	1.336 (0.596, 2.940)	0.862 (0.460, 1.580)	0.609 (0.344, 1.045)	1.419 (0.646, 3.044)
41*	1.256 (1.090, 1.513)	0.865 (0.690, 1.021)	1.461 (1.160, 1.937)	1.257 (1.092, 1.511)	0.852 (0.691, 1.000)	1.476 (1.137, 2.092)
42*	0.768 (0.456, 1.053)	0.774 (0.483, 1.031)	0.993 (0.553, 1.750)	0.698 (0.386, 1.010)	0.715 (0.416, 1.003)	0.974 (0.531, 1.742)
43*	1.235 (1.079, 1.457)	0.890 (0.708, 1.067)	1.391 (1.106, 1.841)	1.236 (1.092, 1.421)	0.894 (0.730, 1.065)	1.384 (1.074, 1.841)

Shown are posterior medians and 95% credible intervals. For topics 1 through 38 analyses used priors that provide information on the sign of the correlation coefficient at the between-study level. Topics 39 through 43 were analyzed using uninformative priors.
*These topics refer to mutually exclusive outcomes.

Appendix Table 6. Estimates with univariate and multivariate meta-analysis using the binomial or multinomial likelihood in topics 44 and 45 (random effects)

Analysis or Topic	$\exp(\theta_1)$	$\exp(\theta_2)$	$\exp(\theta_3)$	$\exp(\theta_1-\theta_2)$	$\exp(\theta_1-\theta_3)$	$\exp(\theta_2-\theta_3)$
Univariate						
44	1.765 (1.053, 2.856)	1.412 (0.902, 2.132)	1.378 (0.902, 2.089)	1.250 (0.646, 2.412)	1.279 (0.663, 2.430)	1.023 (0.555, 1.860)
45	0.973 (0.371, 2.140)	0.979 (0.511, 1.708)	0.698 (0.359, 1.146)	0.992 (0.329, 2.766)	1.404 (0.475, 3.825)	1.405 (0.623, 3.305)
Multivariate						
44	1.578 (0.998, 2.535)	1.441 (0.934, 2.151)	1.492 (0.990, 2.185)	1.099 (0.783, 1.638)	1.074 (0.702, 1.644)	0.975 (0.692, 1.330)
45	0.748 (0.381, 1.494)	0.741 (0.450, 1.128)	0.760 (0.463, 1.116)	1.009 (0.672, 1.575)	0.988 (0.628, 1.650)	0.978 (0.834, 1.178)

Shown are posterior medians and 95% credible intervals. Analyses used mildly informative priors for the between-study correlation matrix.

Appendix Table 7. MSE, coverage and bias for random effects univariate meta-analysis averaging over τ_1^2, τ_2^2 and ρ

K	N	$\exp(\theta_1)$	$\exp(\theta_2)$	MSE for $\hat{\theta}_1$ ($\times 10^{-3}$)	MSE for $\hat{\theta}_2$ ($\times 10^{-3}$)	Bias for $\hat{\theta}_1$ ($\times 10^{-3}$)	Bias for $\hat{\theta}_2$ ($\times 10^{-3}$)	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
10	50	1	1	27.8	27.9	-1.6	0.1	98.7	98.7
10	50	1	1.5	28.0	24.6	-0.3	-6.8	98.6	98.7
10	50	1.5	1	24.9	27.8	-5.7	-0.9	98.6	98.8
10	50	1.5	1.5	24.9	25.2	-5.0	-6.1	98.6	98.7
10	100	1	1	14.2	13.9	-0.4	-0.6	98.5	98.7
10	100	1	1.5	14.2	12.7	0.1	-1.3	98.7	98.4
10	100	1.5	1	12.6	14.4	-2.4	-0.6	98.4	98.4
10	100	1.5	1.5	12.8	12.9	-1.9	-2.1	98.3	98.3
10	500	1	1	2.9	2.9	-0.4	0.1	98.2	98.2
10	500	1	1.5	2.9	2.5	0.3	-0.7	98.4	98.3
10	500	1.5	1	2.6	2.8	-0.4	0.5	98.2	98.3
10	500	1.5	1.5	2.6	2.6	-0.3	-0.2	98.3	98.1
20	50	1	1	13.5	13.6	-0.3	0.7	98.1	97.7
20	50	1	1.5	13.5	12.5	-0.7	-9.4	97.9	97.6
20	50	1.5	1	12.7	13.3	-10.9	0.4	97.4	98.0
20	50	1.5	1.5	12.3	12.4	-7.1	-6.7	97.8	97.6
20	100	1	1	7.0	7.1	0.0	-1.4	97.4	97.4
20	100	1	1.5	7.1	6.3	1.0	-3.0	97.3	97.4
20	100	1.5	1	6.4	7.1	-3.6	1.6	97.3	97.3
20	100	1.5	1.5	6.3	6.3	-2.3	-2.0	97.3	97.3
20	500	1	1	1.4	1.4	0.1	-0.1	97.3	97.2
20	500	1	1.5	1.4	1.3	0.8	-0.6	97.0	97.3
20	500	1.5	1	1.3	1.4	-0.4	0.3	97.3	97.1
20	500	1.5	1.5	1.3	1.3	-0.6	-0.2	97.0	96.9

Appendix Table 8. MSE, coverage and bias for random effects univariate meta-analysis averaging over K, N, θ_1 and θ_2

τ_1^2	τ_2^2	ρ	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
0	0	-0.8	10.4	10.2	-2.1	-0.2	98.0	97.8
0	0	-0.5	10.3	10.8	-2.5	-2.1	98.0	98.0
0	0	-0.2	10.7	10.5	-1.2	-2.4	97.9	97.8
0	0.1	-0.8	10.4	10.8	-2.0	-2.4	98.2	97.7
0	0.1	-0.5	10.7	10.4	-0.8	-0.9	97.9	98.1
0	0.1	-0.2	10.5	10.7	-2.0	-3.2	97.9	97.9
0	0.5	-0.8	10.7	10.8	-0.5	-0.6	97.9	97.9
0	0.5	-0.5	10.5	10.7	-1.6	-2.2	98.0	98.0
0	0.5	-0.2	10.6	10.5	-2.0	-1.8	98.0	98.1
0.1	0	-0.8	10.6	10.7	-2.5	-1.6	98.1	97.8
0.1	0	-0.5	10.5	10.6	-3.0	-0.6	98.0	98.0
0.1	0	-0.2	10.6	10.5	-1.3	-2.7	97.8	97.8
0.1	0.1	-0.8	10.7	10.3	-2.3	-1.4	97.8	97.9
0.1	0.1	-0.5	10.4	10.9	-3.1	-1.7	98.0	97.7
0.1	0.1	-0.2	10.5	10.6	-0.2	-1.9	98.1	97.9
0.1	0.5	-0.8	10.9	10.9	-1.9	-3.3	97.8	97.8
0.1	0.5	-0.5	10.5	10.6	-0.4	-3.3	97.9	97.9
0.1	0.5	-0.2	10.8	10.9	-2.2	-0.6	97.9	97.9
0.5	0	-0.8	10.6	10.6	-3.5	-0.9	97.8	98.0
0.5	0	-0.5	10.5	10.9	-0.6	-1.8	97.9	97.9
0.5	0	-0.2	10.4	10.5	-0.9	-1.8	97.8	98.0
0.5	0.1	-0.8	10.7	10.3	0.5	-2.4	97.9	98.1
0.5	0.1	-0.5	11.0	10.5	-1.9	-1.3	98.0	98.2
0.5	0.1	-0.2	10.7	10.4	-2.0	-1.7	97.9	98.2
0.5	0.5	-0.8	10.8	10.6	-2.4	-0.9	97.9	97.8
0.5	0.5	-0.5	10.6	10.6	-1.7	0.2	98.1	98.0
0.5	0.5	-0.2	10.7	10.2	-3.1	-0.8	97.9	98.2

References

1. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics. Oct 2009;10(4):792-805.

2. Pinheiro JC, Bates DM. Unconstrained parameterizations for variance-covariance matrices. Stat Computing. 1996;6:289-96.